OULUN YLIOPISTO
UNIVERSITY of OULU

DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING

MASTER'S THESIS

FULL-DUPLEX UAV RELAY POSITIONING FOR VEHICULAR NETWORKS

Author	Pouya Pourbaba
Supervisor	Prof. Nandana Rajatheva
Second Examiner	Dr. K. B. Shashika Manosha

April, 2019

Pourbaba P. (2019) Full-duplex UAV relay positioning for vehicular networks. Department of Communications Engineering, University of Oulu, Oulu, Finland. Masterís thesis, 46 p.

ABSTRACT

The unmanned aerial vehicles (UAVs) can be deployed as aerial base stations or wireless relays to enhance the coverage and guarantee the quality of service (QoS) of wireless networks. In this thesis, the positioning of a full-duplex (FD) UAV as a relay to provide coverage for an FD vehicular network is investigated. This problem is solved using two different methods. In both of the methods, the problem is formulated using a predefined set of locations for the UAV. Then this problem is solved for different configurations of the ground users and an optimal location is selected for the UAV to operate at.

In the first approach, given the position of the vehicular users on the ground, a novel algorithm is proposed to find a location for the UAV to satisfy the QoS requirements of the vehicles in the network. The positioning problem is formulated as an ℓ_0 minimization which is non-combinatorial and NP-hard, and finding a globally optimal solution for this problem has exponential complexity. Therefore, the ℓ_0-norm is approximated by the ℓ_1-norm. Simulation results show that by locating the UAV using the proposed algorithm the overall performance of the network increases.

In the second approach, the UAV positioning problem is solved using an MAB framework. In this case, a simple scenario where only one source node is communicating with the relay to transmit its message to the base station is considered. Given the location of the source node and the predefined locations of the UAV, the MAB algorithm can successfully identify the optimal location for the UAV so the system achieves the maximum possible sum rate. The Greedy, ϵ-Greedy, and upper confidence bound (UCB) algorithms are used to solve the problem. The comparison of these algorithms based on their regret values reveals that the UCB algorithm outperforms the performance of the other algorithms. Simulation results show that the UCB algorithm can successfully identify the optimal location for the UAV to maximize the sum rate of the communication links.

Keywords: Full-duplex UAV, Relaying, V2V Communications, Convex Optimization, Reinforcement Learning, Multi-armed Bandit, Upper Confidence Bound

TABLE OF CONTENTS

FOREWORD

The focus of this thesis is the study of unmanned aerial vehicles (UAVs) in wireless communication. The objective is to find an optimal location for a UAV which is being used as a wireless relay between a vehicular network and a base station. This thesis was done in the center for wireless communications (CWC) at the University of Oulu. Here, I would like to appreciate the support and guidance of my supervisor Prof. Premanandana Rajatheva. I also would like to thank Dr. K. B. Shashika Manosha and Mr. Samad Aali whose guidance was very critical in carrying out this work. Moreover, special thanks to my family and my friends for their moral support and all the motivations that they provided me throughout the whole time that I was working on my thesis.

LIST OF ABBREVIATIONS AND NOTATIONS

VANET	Vehicular Ad Hoc Network
ITS	Intelligent Transportation System
IoT	Internet of Things
V2V	Vehicle to Vehicle
V2I	Vehicle to Infrastructure
WAVE	Wireless Access in Vehicular Environments
IEEE	Institute of electrical and electronics engineers
RSU	Road Side Unit
UAV	Unmanned Aerial Vehicle
BS	Base Station
5G	Fifth Generation
4G	Fourth Generation
LTE	Long-Term Evolution
D2D	Device to Device
RFID	Radio Rrequency IDentification
NFC	Near Field Communication
FD	Full-Duplex
SI	Self-Interference
SNR	Signal to Noise Ratio
SINR	Signal to Interference plus Noise Ration
DENMs	Decentralized Environmental Notification Messages
ETSI	European Telecommunications Standards Institute
CAMs	Cooperative Awareness Messages
OBU	Onboard Unit
AU	Application Unit
LoS	Line of Sight
NLoS	Non-Line of Sight
LAP	Low Altitude Platform
HAP	High Altitude Platform
MANET	Mobile Ad Hoc Network
RS	Relay Station
AF	Amplify and Forward ()
DF	Decode and Forward (DF)
CF	Compress and Forward (CF)
QoS	Quality of Service
MAC	Medium Access Control
DSRC	Dedicated Short Range Communications
RLS	Recursive Least Square
GRL	Geometric Reinforcement Learning
GPS	Global Positioning System
MAB	Multi-armed Bandit
UCB	Upper Confidence Bound

$\mathcal{D}$	Set of terrestrial users
g_{rb}	Gain of the link from r to b
g_{sr}	Gain of the link from s to r
g_{rv_1}	Gain of the link between r to v_1
g_{rv_2}	Gain of the link between r to v_2
g_{sv_1}	Gain of the link from s to v_1
g_{sv_2}	Gain of the link from s to v_2
g_{V2V}	Gain of the V2V link
γ_{rb}	SNR of the link from r to b
γ_{sr}	SINR of the link from s to r
γ_{rv_1}	SINR of the link between r to v_1
γ_{rv_2}	SINR of the link between r to v_2
γ_{sv_1}	SINR of the link from s to v_1
γ_{sv_2}	SINR of the link from s to v_2
γ_{V2V}	SINR of the V2V link
p_s	Transmit power of the source node
p_r	Transmit power of the relay
p_v	Transmit power of the vehicles in the V2V link
I_r	Residual of the SI at the relay
I_v	Residual of the SI at the V2V users
$\boldsymbol{L}_r$	Matrix of the predefined locations for the relay
$\boldsymbol{s}_{sr}$	Vector of the received powers at r from s
$\boldsymbol{s}_{rb}$	Vector of the received powers at b from r
$\boldsymbol{s}_{v_1r}$	Vector of the received powers at r from v_1
$\boldsymbol{s}_{v_2r}$	Vector of the received powers at r from v_2
$\boldsymbol{e}$	Unit vector to solve the optimization problem
$\boldsymbol{r}_{sr}$	Vector of the rates for the links between r and s
$\boldsymbol{r}_{rb}$	Vector of the rates for the links between b and r
$\boldsymbol{r}_t$	Vector of the total sum rate

1. INTRODUCTION

Wireless Communications is an inseparable part of the everyday life of mankind. The massive amount of devices that are connected to each other such as smart phones, laptops, wearables, different types of sensors, and vehicles require a secure way of communication. Having so many devices connected to each other through a network and interacting with each other requires an efficient and perfect system which makes the communications between them possible. Therefore, one of the crucial challenges of wireless networks is the ability to support the connectivity of a huge number of users [1]. According to [2], there are two fundamental challenges concerning wireless communication. The first one is the problem of fading due to the time-varying characteristics of channels, obstacles on the communication link, and the distance between the transmitter and the receiver. The second challenge which is the interference on the communication link arises due to the fact that in the wireless communications there is no direct and isolated link between the transmitter and the receiver. Other devices in the area might also be communicating with each other which can impose interference on our transmission link.

The wireless communication has enabled the applications of video transmission, live video streaming sessions from different places with only a mobile phone, and so many other applications to be the reality [1]. These new applications have forced the conventional wireless communication systems which were centered around the transmission of voice with low data rates to shift their focus on the transmission of multimedia with high data rates which requires a better error performance to meet the quality of service (QoS) criteria. However, there is always a trade-off between the two criteria of high bit rates and small error rates [3].

When it comes to the design and implementation of wireless channels, there are several technical issues. First of all, the communication channel is totally unpredictable and random. Hence, the signal propagated through a wireless channel goes through random fluctuations and the signal arrived in the receiver is difficult to be detected. The reason for this phenomena is the concept of multipath fading due to reflection and diffraction. Secondly, the radio spectrum is not an abundant source which makes it an expensive resource. Moreover, since this communication is done through a wireless channel, the signal can be picked by anyone in between the transmitter and the receiver. Therefore, the important concept of security plays a crucial role in this type of communication [1]. Therefore, users must be anonymous to maintain their privacy. At the same time, the wireless users must be authorized to send or receive data to prevent the adversarial actions [4].

Another important issue in a wireless communications system is the ability of the system to track the user and send the desired information to the right user [1]. It is a significant feature in vehicular communications where the mobile users are in high-speed motion. Hence, the vehicular ad hoc network (VANET) which is one of the important applications of vehicular communications started gaining attention. In VANETs, the topology of the network is in a rapid change. Therefore, the network must heal itself and adjust to the new topology and establish a proper route between the transmitter and the receiver. The VANETs enable the users who are traveling in a vehicle to communicate with other users who might be in high-speed motion as well. Additionally, they provide the ability for the vehicles to connect to each other without any middleware which is the backbone of the intelligent transportation system (ITS) [5].

Recently, the concept of vehicle-to-vehicle (V2V) communications have started gaining a great amount of attention. The importance of V2V communications comes into play due to road safety and traffic management tasks. However, the aforementioned concepts impose various challenges on the wireless network management such as the huge number of users that are required to be in communication with each other. For example, the self-driving cars need communication with very low delay constraints and high security. In addition to the V2V, vehicles can communicate with the stationary infrastructure by the road which is called the vehicle to infrastructure (V2I) communication. This stack of communication was standardized by IEEE with the name of wireless access in vehicular environments (WAVE) [6].

Self driving cars require a large number of sensors to be mounted on them, these sensors gather data from the environment around the vehicle such as the speed of the other vehicles, their positions, the direction of their motion, and etc. Thus, vehicles can exchange these information with the other vehicles or the road side units (RSUs) to enable the safety issues such as, detecting the objects in the blind spots, better vision during the night by using the infrared sensors to reduce the fatalities on the road [7].

Recently, using unmanned aerial vehicles (UAVs) in wireless communications have gained a great deal of attention. UAVs can be used as base stations (BSs) in the case of emergencies or catastrophes where the terrestrial facilities are destroyed or cannot provide the required service. Moreover, UAVs can be used as relays in the situations where some improvement to the performance of the existing network is required [8]. An algorithm is proposed in [9] to position a UAV as well as optimize its power to maximize the sum rate of the users. Using UAVs as a wireless communication node either as a relay or a BS can decrease the costs of operation. Moreover, deploying them is simpler and more flexible than the conventional BSs.

Due to the recent developments in both vehicular and aerial networks, many challenging and innovative ideas are being implemented in both areas. However, by combining these two concepts and exploiting their advantages the opportunities that they bring can increase. Vehicular networks are the main part of the ITS which are under intensive research to enable safe and efficient transportation system and UAVs can bring flexible and easily deployable communication nodes. In this work, we are investigating the scenario of vehicular networks cooperating with an aerial wireless node. In particular, this scenario includes a couple of vehicles on the ground which are required to communicate with each other and one of the vehicles sends its messages to a base station (BS). However, this BS is considered to be in a location where there is no direct or indirect link between the two end nodes. Therefore, a UAV is utilized to relay the data between the vehicle and the BS.

One conference paper titled "Full-Duplex UAV Relay Positioning for Vehicular Communications with Underlay V2V Links" from section 3 is accepted for publication at 2019 IEEE Vehicular Technology Conference. Moreover, another paper titled "Multi-Armed Bandit Learning for Full-Duplex UAV Relay Positioning for Vehicular Communications" from section 4 is submitted for publication at International Symposium on Wireless Communication Systems 2019.

The rest of this report is structured as follows. Section 2 presents the background material required for the report. In section 3 we formulate an optimization problem which chooses a location for the UAV out of a finite predefined set of locations. In section 4 we solve a simplified version of the UAV positioning problem with the help of reinforcement learning and section 5 concludes this work.

2. BACKGROUND

2.1. Communication in 5G

The recent applications of wireless communication such as online gaming, video streaming, IoT devices, which rely mostly on the huge amount of data and end users impose a new set of requirements such as high data rates, high-frequency ranges, and low latency compared to previous networking architectures. Decades have passed since the first generation of cellular mobile communications was established, where, the transmission of data was only limited to voice. However, the necessary improvements to the already existing systems were made since that time and the next generations of wireless networks appeared to meet the future needs [10, 11].

The current fourth generation (4G) cellular systems will face difficulties facing the huge amount of data being transmitted via billions of devices. Introducing new applications such as virtual reality, IoT systems, Device to Device (D2D) communications, V2V communications increase the need to improve the existing cellular systems in a way to fulfill the requirements of these applications. For example, IoT systems impose the need for tens of thousands of connected devices per cell with an enormous network traffic demand which 4G is not able to support [10]. The fifth generation (5G) network is the technology that will satisfy the high data rate and low latency requirements [12].

To meet the aforementioned requirements 5G networks must improve the data rates almost to 10 times the data rates of the current long-term evolution (LTE) systems, the bandwidth needs to be increased, a huge number of devices must be connected for a long time, and the latency of the 5G must be almost 10 times less than the 4G LTE. Figure 1 illustrates the schematic diagram of the 5G wireless communication networks [10].

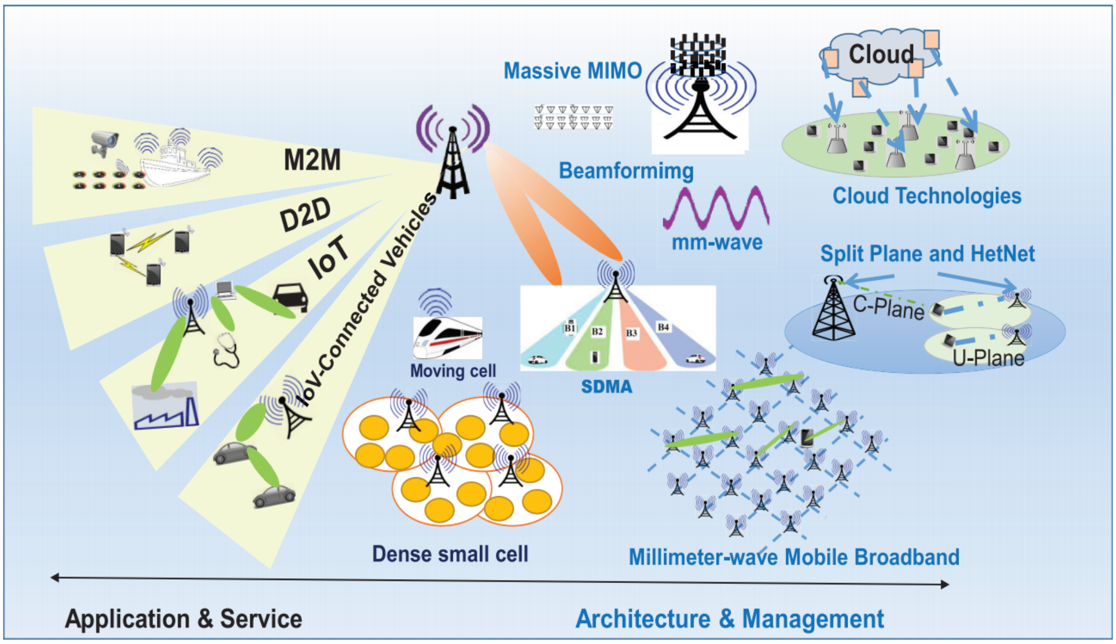

Figure 1: 5G wireless networks [10].

2.2. Internet of Things

The paradigm of IoT consists of the vast amount of devices surrounding us in our everyday life that are connected to a network. This requires an invisible and embedded communication system in the environment. There are one or several sensors mounted on each device and a large amount of data generated by these devices must be transferred to a storage system and the analytical operations are done on them. One of the many definitions of IoT which is based on the radio frequency identification (RFID) group is "The worldwide network of interconnected objects uniquely addressable based on standard communication protocols" (exact words) [13]. The success of RFID technology has played a crucial role in the advent of IoT. However, IoT systems consist of smart objects which are able to carry out sensing of the environment, analyzing the data, and communicating with the other objects [14]. A generic schematic of IoT applications and end users which are categorized based on data is depicted in Figure 2.

The hardware being used to build smart objects are RFIDs, near field communication (NFC), and sensors. RFID is a technology for short-range communication where the device uses an RFID tag to connect to an RFID reader and transmit data. NFC is also a short-range communication technology, where the objects are required to be close to each other to be able to communicate. Sensors are used to measure some particular characteristics of the outside world. When several sensors are cooperating with each other and are connected to a server where they can store the data, they form a so called sensor network [15].

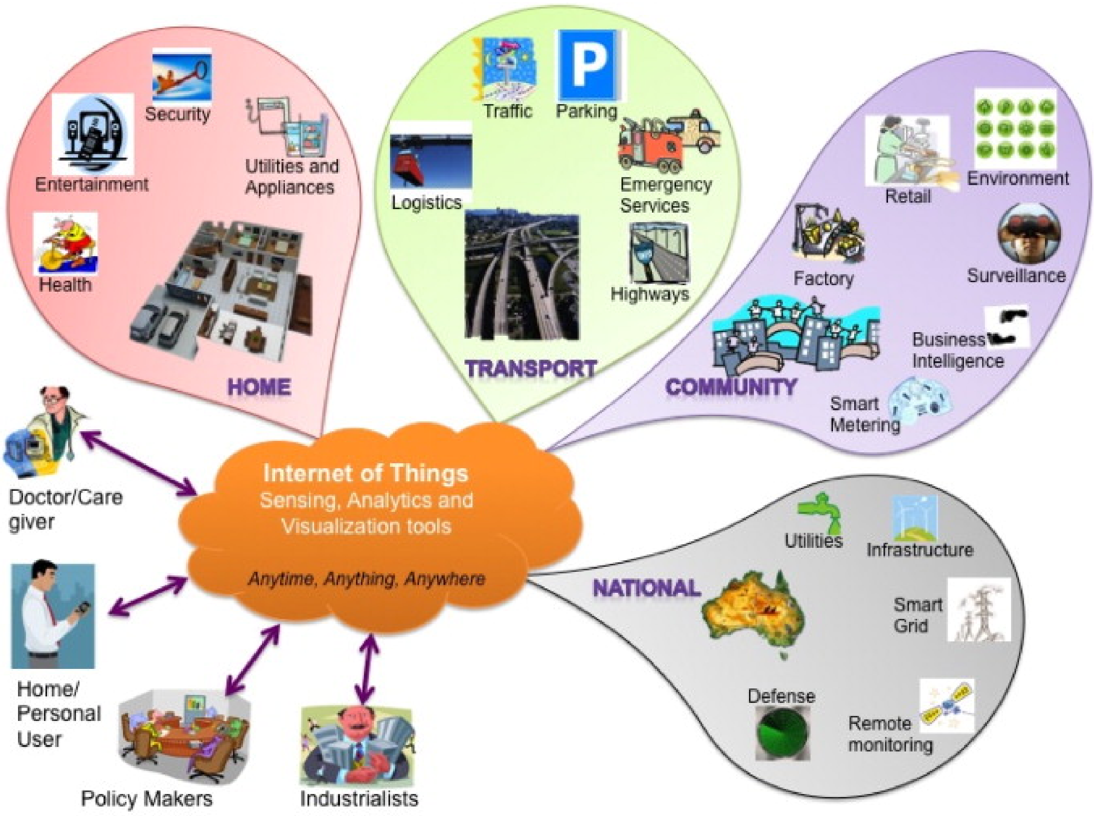

Figure 2: Schematic of the users and applications of IoT [13].

Applying the concepts of IoT in industry, agriculture, and other areas have proved to be beneficial. The key element of the IoT systems is data. The devices with networking capabilities can gather information from the surrounding world and are able to communicate with other devices. This enables these devices to be aware of the context they are in which in turn results in the ability to make proper decisions [16].

2.3. Full-duplex Radios

In the conventional wireless communication systems, the transmit and receive of the data are separate from each other, i.e., the frequency band for transmitting data was not the same as the frequency band for receiving data. Moreover, because of the tendency of the wireless signals to attenuate over distance, the received signal is so much weaker than the transmitted signal. Therefore, identifying the incoming signal in the presence of the outgoing one is considered to be impossible. However, in recent time, the concept of in-band full-duplex (FD) radios has gained tremendous attention both in industry and academia. One of the most important advantages of FD can be the efficiency in the use of bandwidth, no longer we would need two separate channels for uplink and downlink [17–19].

An FD WiFi radio is designed and implemented in [17] where only one single antenna is used to transmit and receive. A novel self-interference (SI) cancellation method is proposed that cancels the interference of the transmitted signal on the received signal. In an FD radio, if the SI is not canceled, it becomes a part of the noise on the received signal and decreases the signal-to-noise ratio (SNR) level of the received signal.

Authors in [18] propose a technique of SI cancellation termed as *antenna cancellation* to implement FD radios. This technique uses two transmit antennas and one receive antenna. The transmit antennas are separated by a distance of $\lambda/2$, where λ is the wavelength of the signal. This separation in space allows the transmitted signals from the two transmit antennas to cancel each other by being added destructively. Therefore, a null position is created where the receive antenna can operate efficiently without the interference from the transmitted signal.

The SI cancellation is the key that enables the implementation of a proper FD radio. Suppose a WiFi signal being transmitted with 20 dBm of transmit power, and a receive antenna located 6-8 inches away from the transmit antenna. There will approximately be -20 dBm of SI on the received signal from the transmitted signal. Moreover, we need to consider almost -93 dBm of noise floor which implies at least a 73 dB of SI cancellation for successful decoding of the received signal. Authors in [20], propose an antenna cancellation approach to overcome the SI phenomenon. The antenna cancellation technique leverages the concept of *signal nulling*, where, two copies of the same signal are added π out of phase to create a destructive add which results in the cancellation of the signals. Figure 3 illustrates both receive antenna cancellation and transmit antenna cancellation methods used in [20]. The work in [21] shows that it is possible to implement an in-band FD communication system using only a single antenna with one carrier.

The use of directional antennas to implement SI cancellation is proposed in [22]. This technique allows a passive suppression of the SI to be achieved, which results in

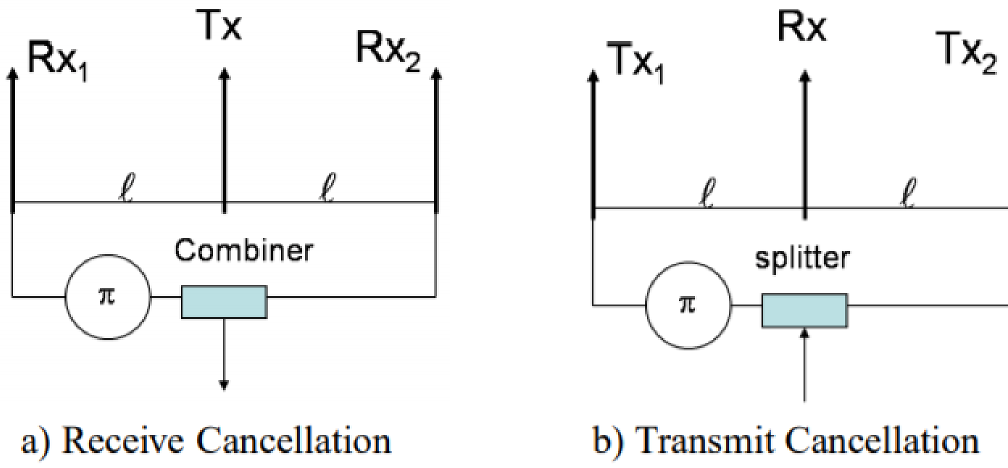

Figure 3: Antenna Cancellation [20].

attenuation of the interference in the electromagnetic level. By setting up directional antennas, we can transmit in a different direction compared to the direction of the received data. Figure 4 illustrates a base station using the directional antennas technique to transmit on downlink while it is receiving in the uplink.

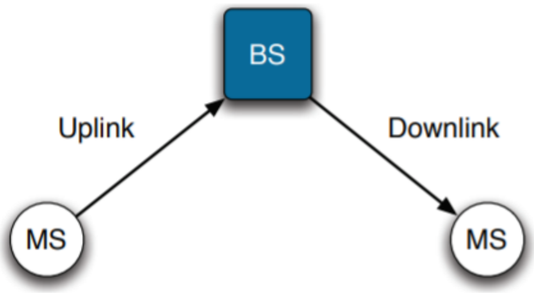

Figure 4: A base station performing a full-duplex communication using directional antennas [22].

2.4. Vehicular communication

Vehicular communication plays a crucial role in providing safety on the roads, traffic management, and environmentally friendly transportation. These applications can operate hand in hand in providing ITS, for example, reducing the number of accidents on the road leads to fewer traffic jams which in turn can decrease the level of environmental damages produced by the vehicles [23, 24]. Enabling these applications to operate require a new set of networking system to support the short-message broadcasting around the vehicles and support the periodic communication of the RSUs with a control system [25]. According to [24], for the vehicles to communicate with other vehicles or the infrastructure around them the VANETs must be utilized. Cellular networks are able to provide the basic data transmission between the passengers of the vehicles, however, the real vehicular communication is possible via VANETs.

There are two main categories of messaging between the vehicles, event-driven and periodic. When an unpredicted situation occurs an event-driven message is sent to the other vehicles. This type of messaging is termed as decentralized environmental notification messages (DENMs) under the European Telecommunications Standards Institute (ETSI). The periodic type of messaging for sharing the current status of the vehicle with its neighboring vehicles, cooperative awareness messages (CAMs) is the term chosen by ETSI to refer to this type of messages [24, 25].

The V2V and V2I communications are done under the WAVE standard. The main components of the WAVE standard are an onboard unit (OBU), an application unit (AU), and RSU. The OBUs are devices on the vehicle enabling the vehicle to communicate with the neighboring environment. The AUs are also located on the vehicle with the task of using the services provided to them from other units. The RSUs are the devices that are located alongside the road and are capable of communicating with the vehicles in their communication range [6].

2.5. Unmanned Aerial Vehicles

Recently, the use of UAVs in wireless communications is rising rapidly. They are utilized in different scenarios, such as aerial base stations, aerial user equipments. A UAV which acts as an aerial base station can provide reliable, low-cost wireless communications. One of the advantages of using an aerial base station is the possibility of having line-of-sight (LoS) links between the BS and the users. Moreover, UAVs come with the mobility feature which enables the UAVs to be deployed easily. In situations where a natural catastrophe has happened and the terrestrial communication infrastructure is out of service or in the hot-spot areas where additional support to cover all the wireless users is a benefit, the UAVs with communication capabilities can be utilized [26].

UAVs can be classified according to two different criteria, the altitude at which the UAV is able to operate and the type of the UAV which identifies the movement features of the UAVs. The UAVs that are deployed in lower altitudes are categorized as low altitude platforms (LAPs) and the UAVs that are able to be deployed in higher altitudes are called high altitude platforms (HAPs). HAPs can be in altitudes higher than 17Km, however, LAPs can be utilized in altitudes of less than a few kilometers. The other criterion for categorizing UAVs is the type of UAV, which can be either fixed-wing or rotary-wing. A fixed-wing UAV needs to be in high-speed motion to be able to stay in the sky. However, a rotary-wing UAV can be in a fixed location in the sky [26]. Figure 5 illustrates the different classes of UAVs and lists their features.

2.6. Wireless Relaying

In a situation where there is a strong shadowing, the deployment of a wireless relay can guarantee a reliable and high capacity wireless transmission. Usage of relays in mobile ad hoc networks (MANETs) not only will overcome the shadowing problems but also will decrease the power consumption of the transmitters [27]. Moreover, having a relay can increase the coverage of a system by enabling the users outside the coverage radius

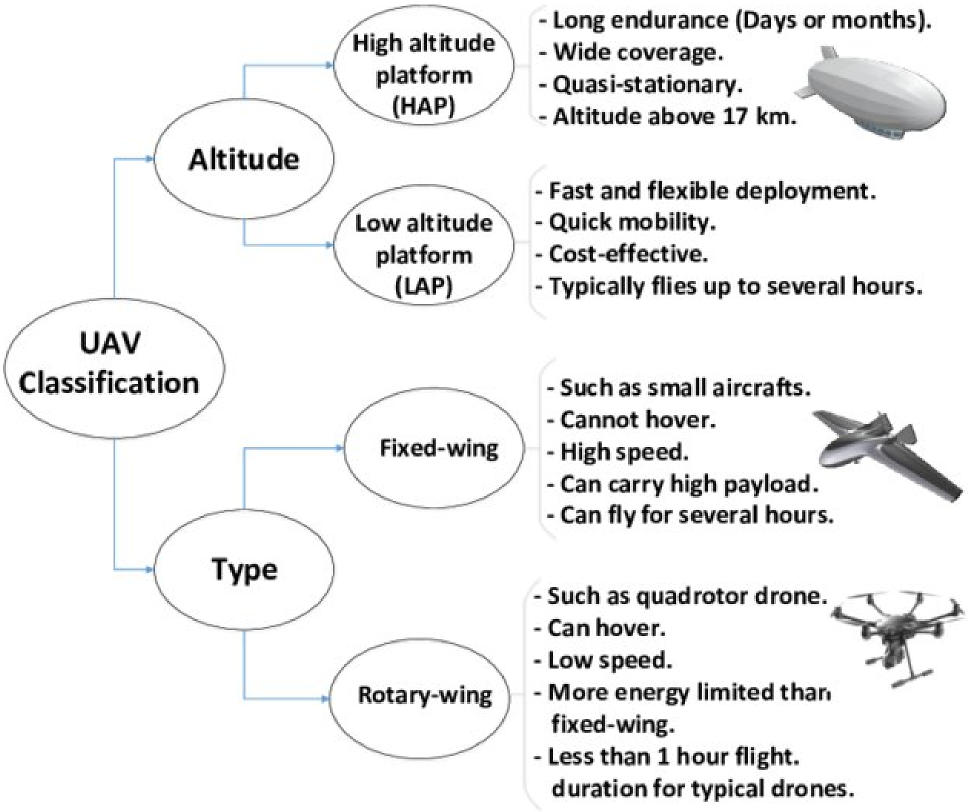

Figure 5: UAV Classification [26].

of a BS to communicate with it. Figure 6 illustrates a network with the deployment of relay stations and the capacity and coverage enhancements provided by them.

There are three general categories for relaying applications: fixed, nomadic, and mobile. Fixed relay stations (RSs) are infrastructures like BSs which are deployed in particular places to support the users in locations exposed to shadowing. However, in hot-spot points which do not have a specific location, the nomadic RSs can be used with lower costs compared to fixed RSs. The mobile RSs are the ones mounted on UAVs, trains, buses, or other means of transportation which can quickly move [28].

Three general methods used by the RSs to implement the process of relaying the messages from a source node to the desired destination node are amplify-and-forward (AF), decode-and-forward (DF), and compress-and-forward (CF). The most basic form of relaying is the AF, in this method the relay only performs the task of amplifying the received signal and transmitting it to the destination. In a DF type of relay, the received signal is demodulated, decoded, encoded again, and modulated again before transmission. The DF technique has a better quality of service (QoS) than the AF, however, it has higher complexity and cost of implementation. The last method is a mixture of both AF and DF which compresses the received signal instead of decoding it [28].

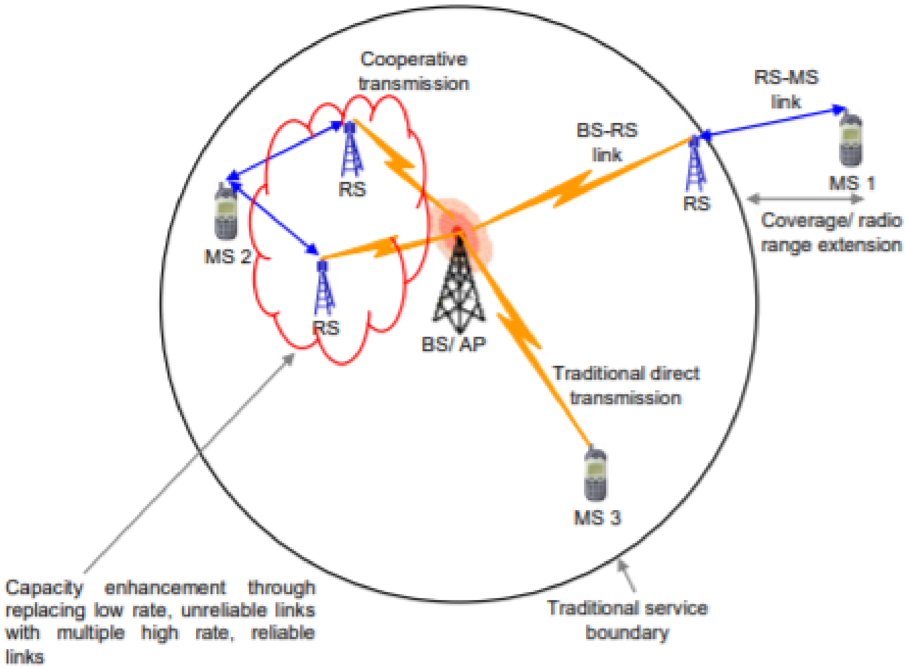

Figure 6: A network with a couple of relay stations [28].

2.7. Learning

Machine learning provides us with algorithms that can learn from data and solve the problems that are not easy to be solved by the conventional programming tools. Some of the tasks that machine learning is capable of doing are classification, regression, and anomaly detection [29]. In other words, with machine learning it is possible to learn a function that maps a set of input variables to a set of output variables. Basically, the target function that is to be learned is unknown and we would like to use the data to learn it. When this function is learned it can be used to predict the output value of the new inputs. However, the learned function will include some error in the predictions. The source of this error can be the lack in the amount of data to be learned from or the lack of descriptive features in the input data. Eq 1 describes the relationship between the input data and the output data and the error,

$$Y = f(X) + e, \tag{1}$$

where Y denotes the output data, f is the target function, the input data is denoted by X and the error is shown as e.

The task of learning is accomplished by using learning algorithms. These algorithms are mathematical models that try to find the best characteristics for the target function and minimize the value of the error to get the optimum predictions [How Machine Learning Algorithms Work (they learn a mapping of input to output)]. There are three general types of machine learning algorithms, which are supervised learning, unsupervised learning, and reinforcement learning.

2.7.1. Supervised Learning

In supervised learning, the data is comprised of a set of input matrix where every row of the matrix includes at least one column (also known as the features) and a vector of desired output associated with the input matrix. by feeding the input data into the mathematical model which has an error function we can compare the results obtained from the model to the actual output data and by minimizing the value of the error function the algorithm can learn to predict the results of new input data points. In supervised learning the output values act like a supervisor to the algorithm. In other words, the algorithm makes some predictions based on some assumptions and then the prediction is compared to the real output value and the amount of error is calculated. Then the algorithm makes some changes in the assumptions in order to minimize the value of the error. This task is repeated till the stopping criteria is met and the error has reached its minimum.

Two groups of supervised learning problems are classification and regression. The type of the output values determines the group which the problem belongs to. When the output values are categorical the problem is considered as classification, whereas, the continuous real-valued outputs result in a regression problem.

2.7.2. Unsupervised Learning

Unsupervised learning algorithms are the set of algorithms that are fed with the input data without any output labels. These algorithms look for patterns in the data and classify them into different clusters based on their similarities. These algorithms are called unsupervised learning because of the fact that there is no label vector to be used in the training process to identify the correct output from the incorrect one. Therefore, the algorithm is on its own to learn the underlaying structure of the data. A clustering problem is considered as un supervised learning problem.

In the case of the supervised and unsupervised learning the input data is usually divided into two sets. The training set and the test set. The training set is used only during the training phase and when the training phase is over, the test set is used to evaluate the model by computing the accuracy of the model. The whole dataset is split into two sets of training data and testing data to have a better evaluation of the model with the data points that it has not seen before. This is termed as generalization and it is a measure of the ability of a machine learning model to produce a correct result for a brand new data point.

2.7.3. Reinforcement Learning

Reinforcement learning is another category of machine learning where the problem is formulated in a manner that there is an agent which is interacting with the world around it which is termed as the environment. In reinforcement learning the agent takes some random actions from an unknown probability distribution and based on that action ends up in a new environment. According to the action and the new environment the agent receives an award. The goal of the agent is to gather the maximum possible amount of award. Taking an action not only can determine the immediate reward but also can influence the future rewards also. Therefore, the agent should have a fair balance between exploring new actions and exploiting the ones that it knows would yield a high reward [30].

A reinforcement learning system is consisted of a policy, a reward signal, a value function, and a model of the environment. A policy determines how the agent would behave in a certain situation. A reward signal is the reward that the agent receives after taking every action. The agent tries to maximize the total reward that it gets in the long run. A value function determines the expected reward the agent can accumulate by being in that state.

3. UAV POSITIONING WITH ℓ_0 MINIMIZATION

Autonomous vehicles are expected to increase the efficiency and safety of future transport systems. Recently, there has been a surge of interest in developing intelligent transport systems (ITS) and addressing the associated problems. For example, the safety is a crucial issue especially in the case of self-driving cars which ITS aims to address by concepts such as vehicle-to-vehicle (V2V) communications and vehicle-to-infrastructure (V2I) communications [6, 31]. Another crucial emerging phenomenon which has attracted a lot of attention is the use of unmanned aerial vehicles (UAVs) as base stations (BSs) in emergencies where it is required to recover the wireless network due to damages from natural disasters such as flood or earthquake [32]. Moreover, the UAVs can be deployed as wireless relays to enhance the communication between wireless devices [33]. However, UAV communications and networking for vehicular networks faces serious challenges such as the high mobility of vehicles and air to ground channel modeling [26, 34, 35]. Moreover, it is necessary to design power efficient methods due to the limited capacity of the batteries in the UAVs.

To address the aforementioned challenges, the authors in [8] propose a new framework for locating and modeling several UAVs in a 3D space, where these UAVs are utilized as aerial BSs to collect data from the IoT devices on the ground. The location of the UAV and the uplink power are determined such that the total transmit power of the devices in the network is minimized subject to their SINR constraints. The deployment of a UAV as an aerial BS which is required to provide wireless communication for a geographical area including device-to-device (D2D) communication network is studied in [35]. It is shown that for different D2D user densities, the UAV can be placed in an optimal height to maximize the system sum-rate and the coverage probability.

Vehicles are expected to be equipped with short-range communication technologies to enable the operative awareness or beaconing where vehicles broadcast their status to the surroundings. Two standards that allow direct V2V communication are IEEE 802.11p and longterm evolution V2V (LTE-V2V) [36]. The performance of the IEEE802.11ad medium access control (MAC) and beamforming mechanism are evaluated in [37], where it is shown that IEEE 802.11ad faces serious challenges and some changes are required in order to be able to satisfy the high-bandwidth requirements of vehicular communications. In [38] full duplex radios are porposed to be used in V2V communications since FD radios can achieve up to two times the rate of a conventional half duplex link [39]. The mmWave communication is proposed as another solution for high bandwidth requirement of connected vehicles in [7]. Autonomous vehicles require a large number of sensors to be mounted on them to get information from the surroundings to model the environment around the vehicle. Because of the huge amount of data that autonomous vehicles generate, very high rates are required to transmit the generated data. However, the existing solutions such as 4G and dedicated short range communications (DSRC) cannot meet the high data rate requirement of the autonomous vehicles. Therefore, novel solutions are required for autonomous vehicular communication systems [7]. With advances in UAV technologies, there has been a surge of interest in using UAVs to address some of the challenges of vehicular communications [26, 40, 41].

A new UAV-assisted data dissemination scheduling strategy in VANETs is proposed in [40] where cooperative data dissemination is used to overcome the limited connection time of the communication links. The authors propose a recursive least square (RLS) algorithm to predict the motion of the vehicles. Moreover, the use of UAVs as assistants for spreading information in vehicular networks is discussed in [26] where the vehicles are grouped in clusters. The UAV is communicating with the head of the cluster, which, decreases the number of links required for the ground users to communicate with each other. Therefore, the interference is reduced and the communication links become more reliable due to the transmit diversity that is introduced by the UAV.

However, to the best of our knowledge, there has been no prior work that investigates the FD UAV relaying in vehicular networks. The main contribution of this work therefore is to propose a novel method for positioning a UAV which is operating as a wireless relay. First, we define a set of locations that the relay can accommodate and operate at. Next, using ℓ_0-norm we formulate a minimization problem for positioning the UAV such that it can satisfy the QoS requirements for the vehicular network. However, ℓ_0 minimization problem is NP-hard and non-combinatorial and finding a globally optimal solution requires exponential complexity. Therefore, we relax all the ℓ_0 functions with their natural ℓ_1-norm approximation and convert it into a convex optimization problem. Additionally, the proposed method will find the optimal height for the UAV to operate efficiently.

The rest of this chapter is organized as follows. Section 3.1 presents the system model describing the air to ground and V2V channel model. In Section 3.2, we formulate the UAV positioning as an ℓ_1 minimization problem and section 3.3 presents the simulation results.

3.1. System Model

Consider a network with a BS, a vehicle as the source node that is communicating with the BS through a UAV relay, and a pair of vehicles which communicate directly with each other through a V2V link. As depicted in Figure 7 the vehicle communicating with the BS is the source node and it is denoted as s, the UAV operating as a relay is shown as r, the BS is denoted as the node b, and the V2V vehicles communicating with each other are shown as v_1 and v_2. We assume that the relay and the V2V vehicles are operating in a full-duplex mode using the same radio resources, i.e., they are able to transmit and receive signals simultaneously on the same frequency band. Furthermore, we assume that the BS is located far apart, and hence, it is not possible to establish a direct link between the BS and the vehicle. Therefore, the relay is required to assist the source node to communicate with the BS. We denote all the ground devices in the system by the set $\mathcal{D} = \{s, b, v_1, v_2\}$. The coordinates of ith device where $i \in \mathcal{D}$ is given by (x_i, y_i, z_i), and the coordinates of the relay are shown as (x_r, y_r, z_r). Furthermore, we assume that the vehicles have GPS functionality and they send their location to the UAV where these locations are stored and updated in a location table and using this table UAV carries out the computations for the optimization problem formulated in Section 3.2 and makes the decision regarding the position that it needs to accommodate. Note that one could also assume that the UAV sends the locations of the vehicles to the BS for solving the optimization problem and finding the optimal location of the UAV [42]. Figure 8 illustrates the coordinate system for this setup, where the middle of the junction in Figure 7 is considered as the origin of the coordinate system.

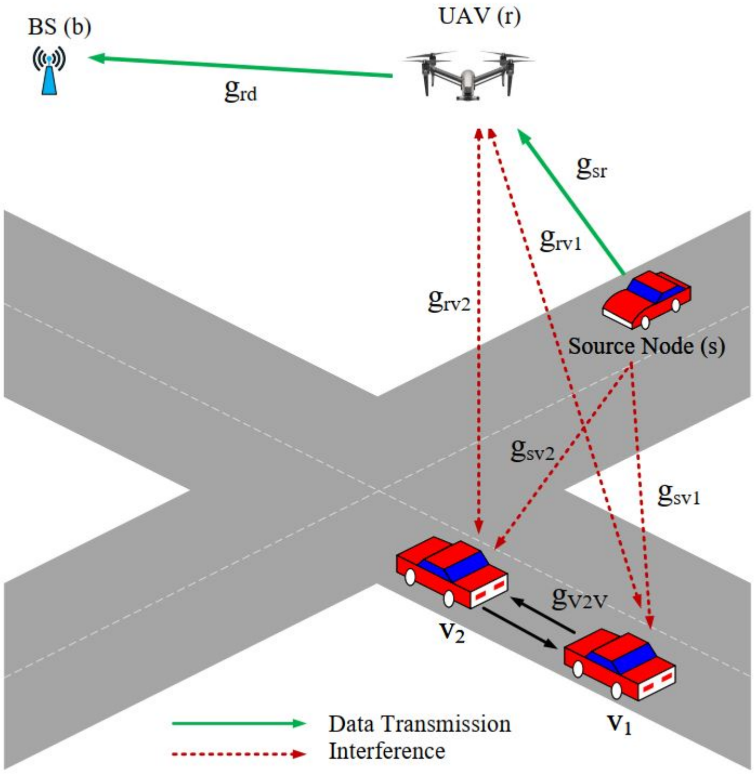

Figure 7: System model.

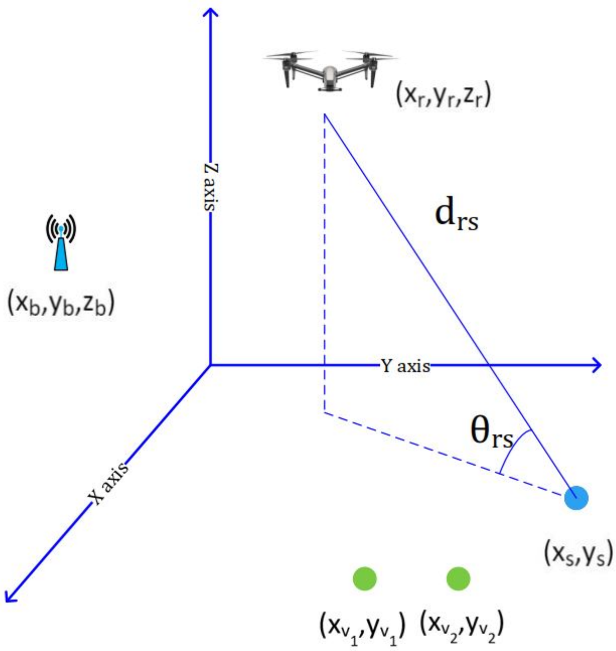

Figure 8: Coordinate system.

3.1.1. V2V Channel Model

The path loss for vehicular communications is considered to follow the dual-slope model [43] which is given by

$$PL(d) = \begin{cases} PL_0 + 10n_1 \log_{10}(d/d_0) + X_\sigma, & \text{if } d_0 \le d \le d_b \\ PL_0 + 10n_1 \log_{10}(d_b/d_0) & \\ +10n_2 \log_{10}(d/d_b) + X_\sigma, & \text{if } d \ge d_b, \end{cases} \tag{2}$$

where d is the distance between the transmitter and the receiver, d_0 is the reference distance, PL_0 is the path loss at the reference distance, X_σ is a zero-mean normally distributed random variable with standard deviation of σ. The notation d_b denotes the breakpoint distance where the first Fresnel zone touches the ground, n_1 is the path loss exponent until the distance d_b, and n_2 is the path loss exponent for the distances after d_b. The breakpoint distance is defined as

$$d_b = \frac{4h_{TX}h_{RX} - \frac{\lambda^2}{4}}{\lambda}, \tag{3}$$

where h_{TX} and h_{RX} are the transmitter and the receiver heights, respectively, and λ is the wavelength.

3.1.2. Air to Ground Channel Model

In the air-to-ground channel model there are two main groups of signals received in the receiver, the first group is the line of sight (LoS) and the second group is the non-line of sight (NLoS) [32]. The occurrence probabilities of the LoS and NLoS links are a function of the environment and the elevation angle between the UAV and the ground user. The parameters defining these probabilities are the average number of the buildings per square kilometer, distribution of the heights of the buildings, and the ratio of the area with buildings to the whole area. The path loss for the LoS and NLoS components can be calculated as a function of the distance between the relay and the ground devices. The following are the path loss equations for the LoS and NLoS links [44]:

$$L_{LoS} = \eta_{LoS}(\frac{4\pi f_c d_i}{c})^\mu, \tag{4}$$

$$L_{NLoS} = \eta_{NLoS}(\frac{4\pi f_c d_i}{c})^\mu, \tag{5}$$

where η_{NLoS} and η_{LoS} ($\eta_{NLoS} > \eta_{LoS} > 1$) are the excessive path loss coefficients which are defined according to the propagation group and the physical environment [44], c is the speed of the light, f_c is the carrier frequency, μ is the path loss exponent, and d_i is the distance between the relay and the user $i \in D$ on the ground and is calculated by

$$d_i = \sqrt{(x_r - x_i)^2 + (y_r - y_i)^2 + (z_r - z_i)^2}. \tag{6}$$

As discussed before, the LoS and NLoS links have their own probabilities of occurrence which depend on the environment characteristics and according to [44] they can be expressed as

$$P(LoS) = \frac{1}{1 + \alpha \exp(-\beta[\frac{180}{\pi}\theta_i - \alpha])}, \tag{7}$$

$$P(NLoS) = 1 - P(LoS), \tag{8}$$

where α and β are constants depending on the type of the environment, θ_i is the elevation angle between the UAV and the ground user $i \in \mathcal{D}$, $\theta_i = \frac{180}{\pi} \times \arcsin(\frac{h_i}{d_i})$, where h_i is the vertical distance between the relay and the node $i \in \mathcal{D}$. The following equation denotes the average path loss between the ground user i and the relay [45]:

$$L = P(LoS) \times \eta_{LoS}(\frac{4\pi f_c d_i}{c})^{\mu} + P(NLoS) \times \eta_{NLoS}(\frac{4\pi f_c d_i}{c})^{\mu}. \tag{9}$$

The SINR of the link between the source and the relay is given by

$$\gamma_{sr} = \frac{p_s g_{sr}}{N_0 + I_r + \sum_{j=1}^{2} p_{v_j} g_{rv_j}}, \tag{10}$$

where N_0 is the additive white Gaussian noise, p_s is the transmit power of the source, g_{sr} is the channel gain between the source and the relay, I_r is the residual of the self-interference (SI) [34]. The residual of SI is defined as $I_r = \delta p_r$ where δ depends on the SI cancellation method. The term $\sum_{j=1}^{2} p_{v_j} g_{rv_j}$ is the total interference from the full-duplex connected vehicles, p_{v_j} is the transmit power of the j^{th} vehicle in the full-duplex link and g_{rv_j} is the channel gain between jth vehicle and the relay. The SINR of the FD V2V link can be written as

$$\gamma_{v_j} = \frac{p_{v_k} g_{V2V}}{N_0 + I_{v_j} + p_s g_{sv_j} + p_r g_{rv_j}} \quad j, k \in \{1, 2\}, j \neq k, \tag{11}$$

where p_{v_k} is the power of the signal transmitted from the vehicle v_k, g_{V2V} is the channel gain between the two vehicles, I_{v_j} is the residual of the SI, $p_s g_{sv_j}$ is the interference coming from the source node to the jth V2V user, $p_r g_{rv_j}$ is the interference coming from the relay to the jth V2V user. The SNR of the link between the r and b is calculated by

$$\gamma_{rb} = \frac{p_r g_{rb}}{N_0}, \tag{12}$$

where p_r is the transmit power of the relay, and g_{rb} is the channel gain between the r and the b. Since the interference generated by other links can be neglected as we assumed the destination node is located far apart from other links for the link between r and b we have SNR instead of SINR.

3.2. Problem formulation and solution approach

In this section, we first formulate the problem of UAV placement. This problem is NP-hard and nonconvex. Therefore, we approximate the original problem with a convex optimization problem and propose a suboptimal method to solve the approximated problem.

3.2.1. Problem formulation

Let us first define the following set of notations. Let $l_i \in \mathbb{R}^3$ be the ith location that the relay can operate at, where the first, second, and the third elements of l_i are x, y, and z coordinates, respectively. This matrix is expressed as

$$\boldsymbol{L}_r = \begin{bmatrix} x_{r_1} & y_{r_1} & z_{r_1} \\ x_{r_2} & y_{r_2} & z_{r_2} \\ \vdots & \vdots & \vdots \\ x_{r_l} & y_{r_l} & z_{r_l} \end{bmatrix}. \tag{13}$$

Next, we define the received signal power $\boldsymbol{s}_{sr} \in \mathbb{R}^L$ at each location of the r from s. This vector can be expressed as

$$\boldsymbol{s}_{sr} = p_s \boldsymbol{g}_{sr}. \tag{14}$$

where, p_s is the transmit power of the s and $\boldsymbol{g}_{sr} \in \mathbb{R}^L$ is channel gain vector for the links between s and each of the predefined locations for r. Similarly, we define the interference signal powers received at each location for the relay from the V2V vehicles as $\boldsymbol{s}_{v_1r} \in \mathbb{R}^L$ and $\boldsymbol{s}_{v_2r} \in \mathbb{R}^L$. These vectors are expressed as

$$\boldsymbol{s}_{v_1r} = p_{v_1} \boldsymbol{g}_{v_1r}, \tag{15}$$

$$\boldsymbol{s}_{v_2r} = p_{v_2} \boldsymbol{g}_{v_2r}. \tag{16}$$

where, p_{v_1} and p_{v_2} are the transmit powers of the first V2V vehicle and the second V2V vehicle, respectively. Furthermore, vectors $\boldsymbol{g}_{v_1r} \in \mathbb{R}^L$ and $\boldsymbol{g}_{v_2r} \in \mathbb{R}^L$ are the vectors of the channel gains between the V2V vehicles and r in each of the predefined locations. The received signal power at b from each predefined location of r is denoted as $\boldsymbol{s}_{rb} \in \mathbb{R}^L$ and can be expressed as

$$\boldsymbol{s}_{rb} = p_r \boldsymbol{g}_{rb}. \tag{17}$$

where, the notation p_r is the transmit power of the node r and $\boldsymbol{g}_{rb} \in \mathbb{R}^L$ is the vector of the gains for the links between the locations of r and the b. Finally, we define the interference signal powers received at the V2V vehicles from r by $\boldsymbol{s}_{rv_1} \in \mathbb{R}^L$ and $\boldsymbol{s}_{rv_2} \in \mathbb{R}^L$. These vectors are expressed as

$$\boldsymbol{s}_{rv_1} = p_r \boldsymbol{g}_{rv_1}, \tag{18}$$
$$\boldsymbol{s}_{rv_2} = p_r \boldsymbol{g}_{rv_2}. \tag{19}$$

where, vectors $\boldsymbol{g}_{rv_1} \in \mathbb{R}^L$ and $\boldsymbol{g}_{rv_2} \in \mathbb{R}^L$ are the vectors of the channel gains between the r in each of the predefined locations and V2V vehicles.

In order to be able to choose the best location from the set $\mathcal{L}$ we define a vector $\boldsymbol{e} \in \mathbb{R}^L$ whose entries must be all null except one entry which must be equal to one. The index of the nonzero entry in $\boldsymbol{e}$ demands that the set of coordinates with the same index from the location matrix $\boldsymbol{L}$ must be chosen for the relay to be located at. The vector $\boldsymbol{e}$ is expressed as

$$\boldsymbol{e}^T = \begin{bmatrix} 0 \dots 1 \dots 0 \end{bmatrix}. \tag{20}$$

By using notation $\boldsymbol{e}$, we now rewrite the SINR expressions defined in (10), (11), and (12). The SINR of the link from s to r is expressed as

$$\gamma_{sr} = \frac{\boldsymbol{e}^T \boldsymbol{s}_{sr}}{N_0 + I_r + \sum_{j=1}^{2} \boldsymbol{e}^T \boldsymbol{s}_{rv_j}}. \tag{21}$$

Algorithm 1 UAV positioning algorithm

1: For a given topology: set SINR thresholds
2: Find SINRs for each location of r
3: Approximate problem (24) by (25)
4: Solve (25) and find $\boldsymbol{e}$
5: Find the index of the maximum value in $\boldsymbol{e}$, and locate the UAV

The SINR for each of the V2V links is calculated as

$$\gamma_{v_j} = \frac{p_{v_k} g_{V2V}}{N_0 + I_{v_j} + p_s g_{sv_j} + \boldsymbol{e}^T \boldsymbol{s}_{rv_{ji}}} \quad j,k \in \{1,2\}, j \neq k. \tag{22}$$

Finally, we can calculate the SNR for the link from the relay to the destination:

$$\gamma_{rd} = \frac{\boldsymbol{e}^T \boldsymbol{s}_{rd}}{N_0}. \tag{23}$$

Our goal is to find a location for the UAV to operate (from a given set of locations $\mathcal{L}$) such that the QoS of the V2V link, source to relay link, and relay to destination link is guaranteed. The QoS of these links can be guaranteed when the SINR of each link is greater than a predefined threshold. Hence, this design problem can be formulated as the following feasibility problem

$$\begin{aligned}
&\text{minimize} && 0 \\
&\text{subject to} && \frac{\boldsymbol{e}^T \boldsymbol{s}_{sr}}{N_0 + I_r + \sum_{i=1}^{2} \boldsymbol{e}^T \boldsymbol{s}_{rv_i}} \geq \gamma_1 && \text{(24a)} \\
& && \frac{p_{v_j} g_{V2V}}{N_0 + I_{v_i} + p_s \boldsymbol{g}_{sv_i} + \boldsymbol{e}^T \boldsymbol{s}_{rv_i}} \geq \gamma_2,\ i,j \in \{1,2\}, i \neq j && \text{(24b)} \\
& && \frac{\boldsymbol{e}^T \boldsymbol{s}_{rd}}{N_0} \geq \gamma_3 && \text{(24c)} \\
& && \|\boldsymbol{e}\|_0 = 1 && \text{(24d)} \\
& && e_k \in \{0,1\}, \qquad k = 1, ..., L && \text{(24e)}
\end{aligned}$$

where the variable is $\boldsymbol{e}$. Problem (24) is non-combinatorial and NP-hard, and it requires exponential complexity to find a global optimal solution [46]. Therefore, we have to rely on suboptimal methods to find an approximate solution to problem (24).

3.2.2. Solution approach

In the following, we approximate problem (24) as a convex optimization problem. A natural approximation of ℓ_0 is its ℓ_1-norm function. Hence, by replacing all the ℓ_0

Table 1: Environment parameters for A2G channel model.

Environment	η_{LoS}	η_{NLoS}	α	β
Suburban	0.1	21	5.0188	0.3511
urban	1	20	9.6101	0.1592
Dense urban	1.6	23	11.9480	0.1359
High rise urban	2.3	34	27.1562	0.1225

functions with their ℓ_1-norm functions, we can write the approximated problem of (24) as follows

$$
\begin{aligned}
\text{minimize} \quad & 0 \\
\text{subject to} \quad & \frac{\boldsymbol{e}^T \boldsymbol{s}_{sr}}{N_0 + I_r + \sum_{i=1}^{2} \boldsymbol{e}^T \boldsymbol{s}_{rv_i}} \geq \gamma_1 && \text{(25a)} \\
& \frac{p_{v_j} g_{V2V}}{N_0 + I_{v_i} + p_s \boldsymbol{g}_{sv_i} + \boldsymbol{e}^T \boldsymbol{s}_{rv_i}} \geq \gamma_2, \; i, j \in \{1, 2\}, i \neq j && \text{(25b)} \\
& \frac{\boldsymbol{e}^T \boldsymbol{s}_{rd}}{N_0} \geq \gamma_3 && \text{(25c)} \\
& \|\boldsymbol{e}\|_1 \leq 1 && \text{(25d)} \\
& 0 \leq e_k \leq 1, \qquad k = 1, ..., L && \text{(25e)}
\end{aligned}
$$

where the optimization variable is $\boldsymbol{e}$. Note that the binary constraint in (24e) has been relaxed by introducing constraint (25e) in problem (25). This is a convex optimization problem and we can use any standard CVX solver to solve this problem. The proposed algorithm for solving the relay positioning problem is summarized in Algorithm 1.

3.3. Numerical Results

We consider a cross-road in which the source node and the V2V linked vehicles are located. Above this cross-road, we consider a square 114 m $\times$ 144 m area with 400 predefined locations for the relay. Locations are placed with 6m distance from each other. We consider UAV communications in different environments with the carrier frequency of 2GHz. The environment parameters are presented in Table 1 [32]. The BS is located at the coordinates of (1000,1000) and the locations of the vehicles are randomly generated. We set the vehicles to be on either of the streets with a length of 1 km. The distance between the V2V link is set to be 40 m. For obtaining each of the points on the figures, we have run 500 simulations with different vehicular configurations. Each number on the y axis presents the number of times that the problem is solved out of 500 times. In other words, y axis shows how many times a proper location is found for the relay out of 500 times. Table 2 presents the simulation parameters.

Table 2: Simulation parameters.

Description	Value
V2V transmit power (P_{v_1}, P_{v_2})	0-1 - 0.4 mW
Source transmit power (P_s)	0.5 mW
Relay transmit power (P_r)	0.5 mW
Carrier frequency (f_c)	2 MHz
Bandwidth (BW)	1 KHz
Number of the locations of the relay (l)	400
Reference path loss (PL_0)	63.9 dBm
Reference distance (d_0)	10 m
Break-point distance (d_b)	161 m
Path loss exponent (n_1)	1.81
Path loss exponent (n_2)	2.85
Reference distance (d_0)	10 m
Noise power spectral density (N_0)	-170 dBm
BS antenna height (h_b)	30 m

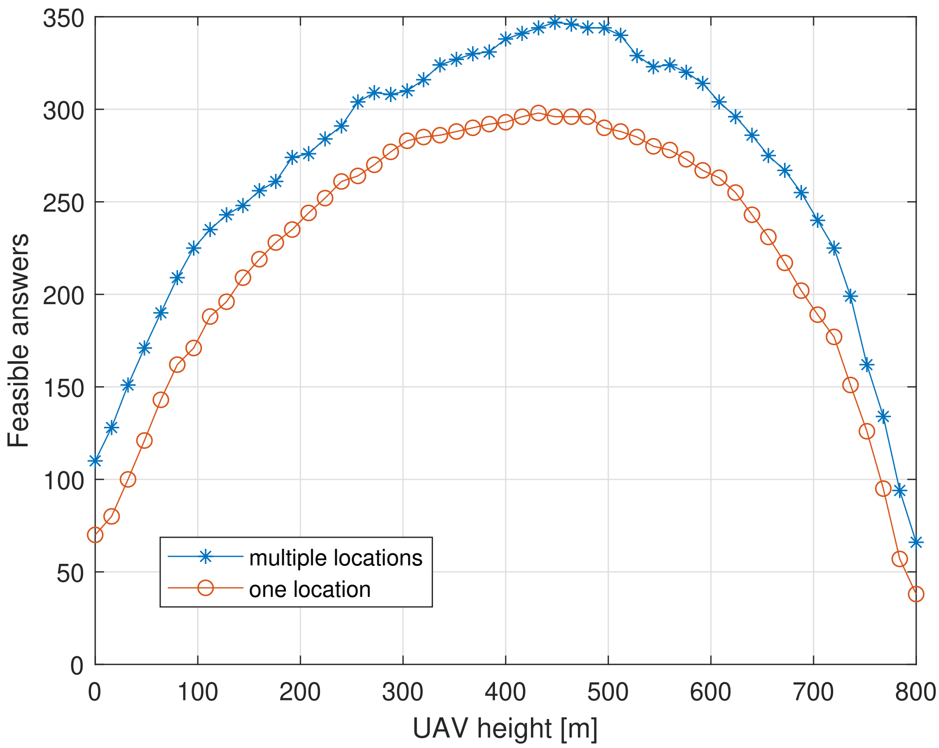

Figure 9: Number of feasible answers for a static UAV with one location compared to multiple locations.

Figure 9 shows the number of of UAV positions meeting SINR threshold for the ℓ_1 minimization problem compared to the number of the times that a UAV in a fixed location can satisfy the SINR requirement of the links in the system. The number of feasible answers of the problem increases by implementing the predefined locations

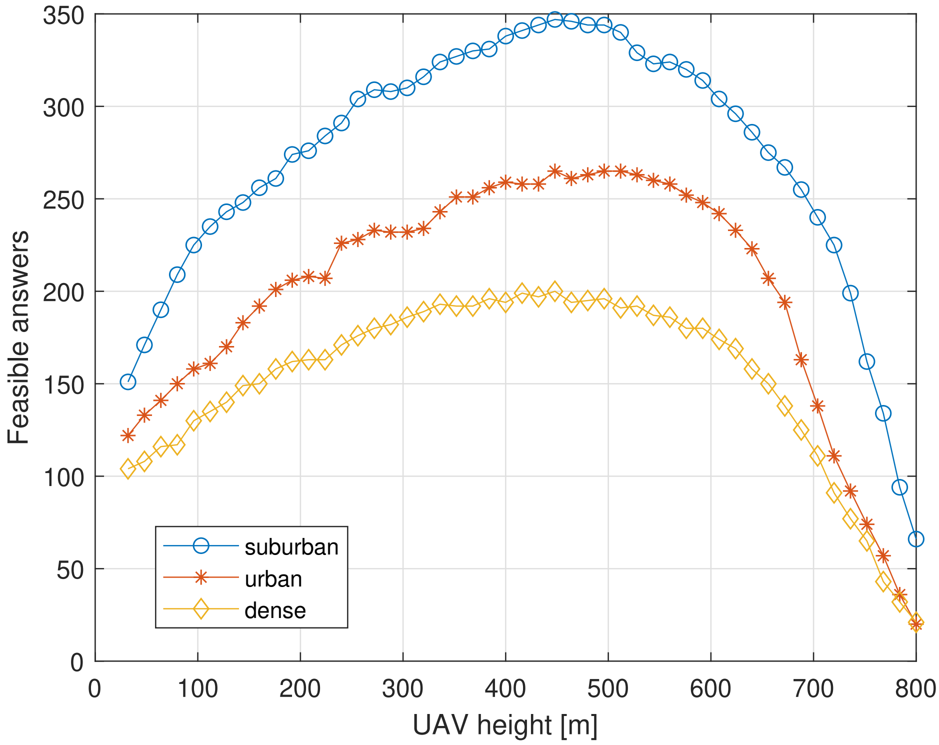

Figure 10: Number of feasible answers in different environments.

for the relay due to the position options that the relay is provided with. These different locations offer a wide range of choices for the relay to choose to operate at. Since, each predefined location is assigned a set of SINRs values, the possibility that the relay can find a position that the SINRs set would satisfy the constraints increases. However, this does not happen when the UAV has a fixed location, and therefore, our proposed algorithm performs better in satisfying the SINR requirements of the system.

Figure 10 shows the effect of different environments on the system where the same set of vehicle locations with an SINR constraint of 1dB is considered. As it is depicted in Figure 10, in a suburban area the number of feasible answers to the problem are higher than an urban area and the feasible answers in an urban area are higher than that of a dense urban area. Moreover, there is a pattern in the results of all the environments and that is due to the fact that by increasing the altitude of the UAV the possibility of having a LoS link increases and we get more feasible answers. However, at some point this effect stops and the number of the feasible answers decrease due to the increase in the distance between the relay and the ground user which leads to a high pathloss.

Figure 11 shows the number of feasible answers for the problem in an urban environment for different SINR constraints for the V2V link. The SINR constraint for the relay links are set to be 1 dB, however, we change the V2V link SINR threshold from 1 dB to 4 dB. As shown in Figure 11 the number of feasible answers decrease by increasing the SINR threshold.

Figure 12 shows the number of feasible answers for the problem in an urban environment while the transmit power of the source node and the relay are fixed at 0.5 mW but the transmit power of the V2V vehicles are changed from 0.1 mW to 0.4 mW. As shown in Figure 12, the number of UAV positions meeting SINR threshold decreases

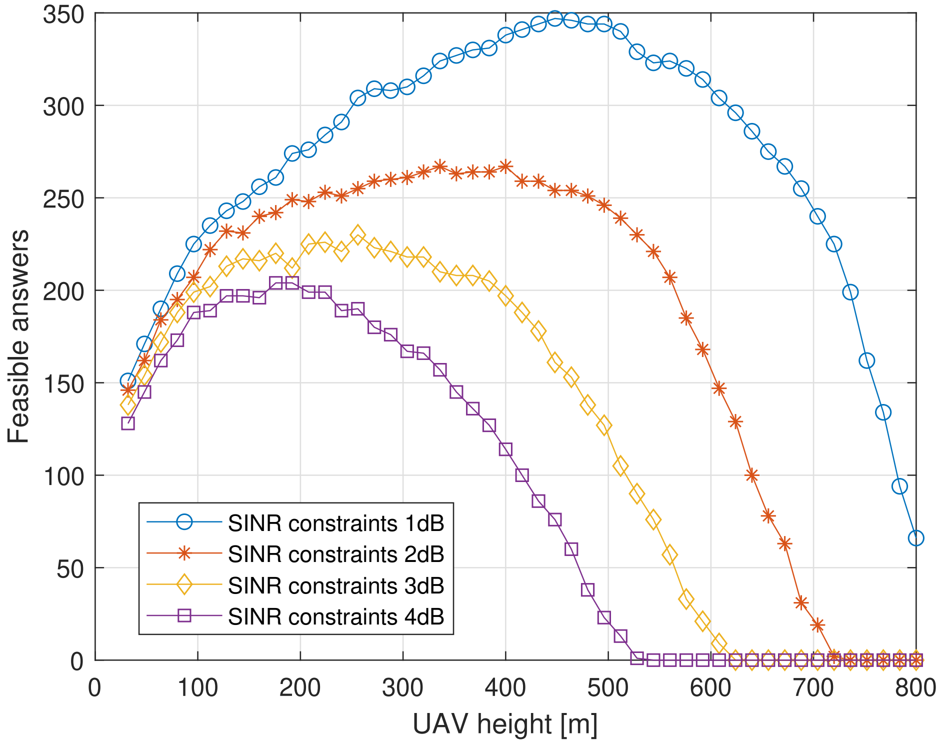

Figure 11: Number of feasible answers for different SINR constraints in a suburban environment.

as the V2V transmit power increases due to the excessive interference from the V2V on the transmission link from s to r. Furthermore, by increasing the transmit power of the s and r the number of feasible answers increase as SINR of the link from s to r improves.

Figure 13 depicts the average sum rate for all the links in the system. Similar to Figure 10 the curve has a rise in its value as the height of the UAV increases which is due to the increase in the probability of LoS for the links from s to r and from r to b. However, this increment stops at the heights around 500 m and the sum rate undergoes a reduction in its value due to the huge path loss for the links because of the high distance between the nodes. Moreover, the point where the system reaches the maximum sum rate is marked with an orange circle.

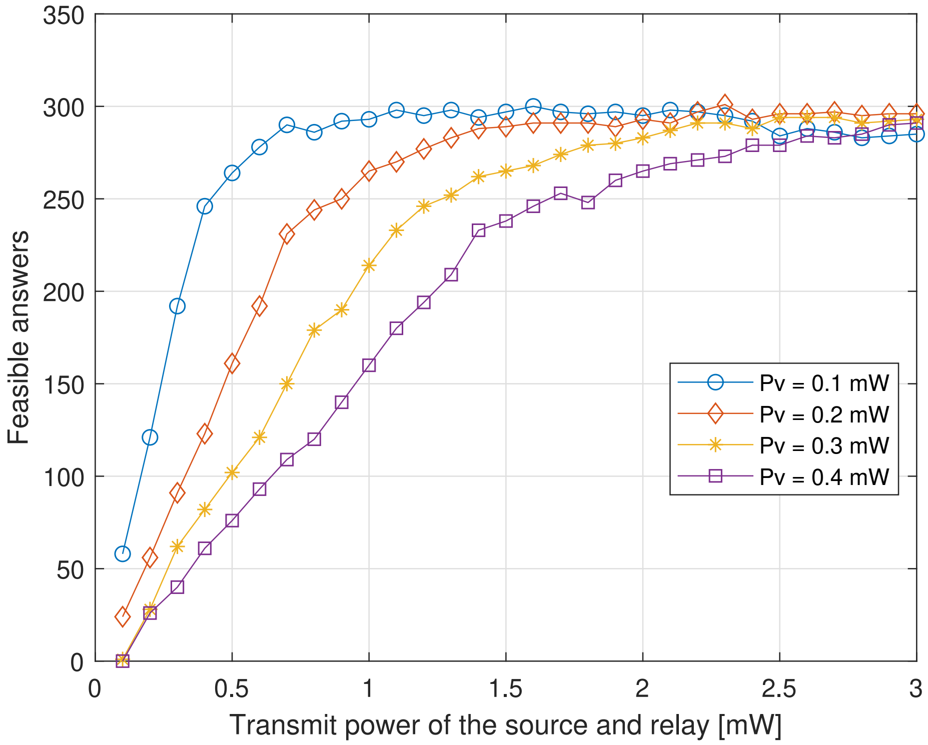

Figure 12: Number of feasible answers for different V2V transmit powers.

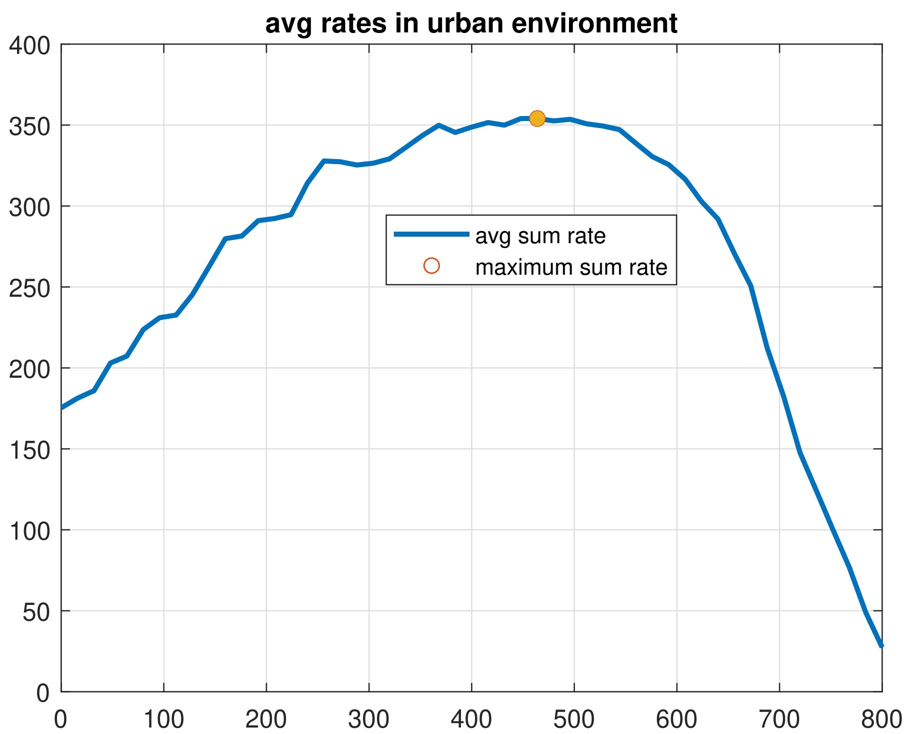

Figure 13: Average sum rate.

4. UAV POSITIONING WITH MACHINE LEARNING

In recent studies the use of reinforcement learning to solve the wireless communication-related problems has increased. The wireless communications related applications which are modeled as a reinforcement learning problem are network selection problems of heterogeneous networks, channel sensing, energy harvesting, and so on [47]. In [48] the reinforcement learning is utilized to allocate the sufficient amount of recourses to the V2V link which reuses the spectrum of the uplink between a vehicle and the BS. The V2V link itself is considered as an agent which decides its own transmission power and finds the optimal sub-band to satisfy the V2V constraints. The authors in [49] use reinforcement learning to transmit delay-sensitive data efficiently over a fading channel. In another work, the UAV is used as a relay to receive the data from a vehicle and transmit it to a roadside unit (RSU) with a better transmission condition than the nearby roadside unit. The UAV acts as the agent and based on the information that it gets from the environment it decides whether or not to relay its message to another RSU [50].

A new algorithm for planning the path of UAVs is proposed in [51], this algorithm is called Geometric Reinforcement Learning (GRL). This algorithm provides a basic and efficient method to plan the path of UAVs. It enables the possibility of assessing the path by its length and risk. In another work, authors have used reinforcement learning to plan the paths of UAVs in a cellular-connected network [52]. The goal of the agents in this work is to maximize the energy efficiency and minimize the latency and the interference generated from the ground users.

4.1. System Model

Consider a vehicular system in which a vehicle is required to communicate with a BS. We consider that because of some geographical conditions or high shadowing on this communication link, the communication link will be in deep fade and transmissions will fail. Therefore, a UAV is used as a relay to provide connectivity between the vehicle and the BS. The vehicle is denoted by v, the UAV operating as the relay is referred to by the letter r, and the letter b is used to the BS. The relay is assumed to be communicating in a full-duplex (FD) manner where it receives and transmits data simultaneously on the same frequency band. The set $\mathcal{D} = \{v, b\}$ which includes the users on the ground is defined to simplify the mathematical equations. The locations of v and b are given by $(x_i, y_i, z_i), i \in \mathcal{D}$. The predefined locations of the relay are defined in the form of a matrix $\boldsymbol{L}_r$ where each row $l_j \in \mathbb{R}^3$ of $\boldsymbol{L}_r$ represents jth location with three columns for the x, y, and z coordinates. The matrix of locations for the relay can be written as

$$\boldsymbol{L}_r = \begin{bmatrix} x_{r_1} & y_{r_1} & z_{r_1} \\ x_{r_2} & y_{r_2} & z_{r_2} \\ \vdots & \vdots & \vdots \\ x_{r_l} & y_{r_l} & z_{r_l} \end{bmatrix}. \tag{26}$$

Figure 14 shows the system model considered for this section.

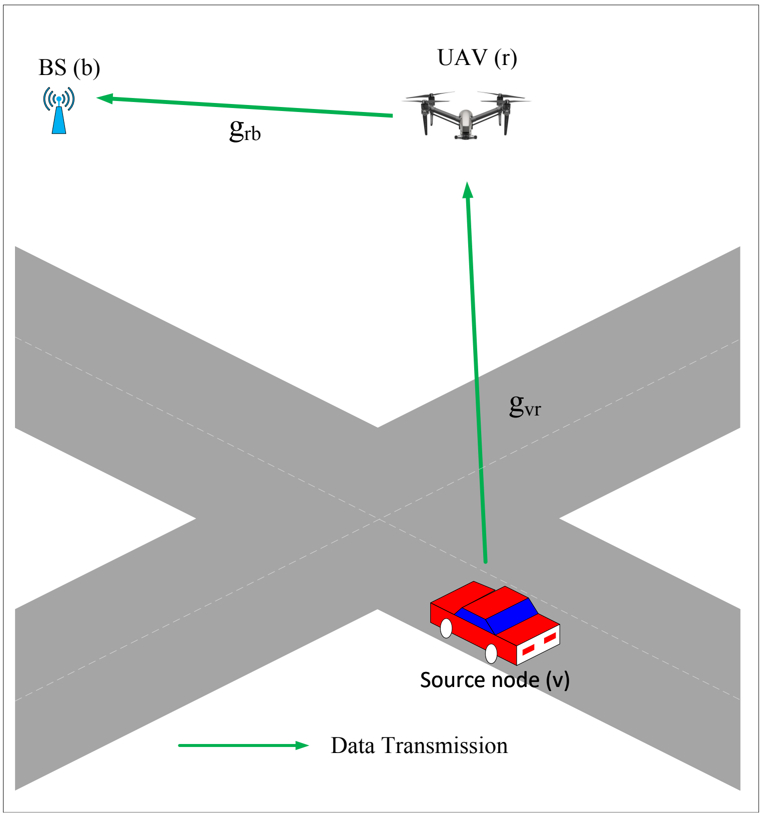

Figure 14: System model 2.

4.2. Multi-armed Bandit

MABs are a form of reinforcement learning where there is a set of available arms (actions) for an agent to select from. When an arm A_t is selected, it generates a reward R_t from a probability distribution which is not known to the agent. The objective of the agent is to maximize the expected total reward. Since the agent does not know the distribution from which the rewards of each arm are drawn it needs a strategy to compensate for the lack of information to achieve its goal [30]. The agent only observes the reward of the arm that it has played. Therefore, the agent can calculate an estimation of the value $Q_t(a)$ for action a before selecting it. The estimation of the action value prior to time t is given by:

$$Q_t(a) = \frac{\text{sums of rewards when action } a \text{ is taken prior to } t}{\text{number of times action } a \text{ taken prior to } t}. \tag{27}$$

The agent can play the arm with the highest value for Q_t, which is known as the greedy action selection method. This method leads to exploiting the arm with the highest estimation for the action value without exploring any other arm. The greedy method only increases the reward at the current time. However, the objective of the agent is to increase the cumulative reward. Therefore, the agent is required to have a reasonable trade-off between exploitation and exploration.

To overcome the challenge of exploration and exploitation the upper confidence bound (UCB) algorithm can be used. Therefore, the UCB algorithm monitors the potential of the non-greedy actions to be the optimal action instead of exploring the

Algorithm 2 UCB Algorithm

Input: τ (horizon), $\mathcal{A}$ (arms)

Play each arm (action) a once
Observe the rewards of each arm r_a
Set $k_a = 1, \forall a \in \mathcal{A}$
Set $\hat{\mu_a} = \frac{r_a}{k_a}$
for $t = |\mathcal{A}|$ to τ **do**
    Play arm $\hat{a} = \arg\max_a \left(\hat{\mu_a} + \sqrt{\frac{2\ln(t)}{k_a}}\right)$
    Observe reward r
    $r_{\hat{a}} = r_{\hat{a}} + r$
    $k_{\hat{a}} = k_{\hat{a}} + r$
    Update $\hat{\mu_a} = \frac{r_a}{k_a}$
end for

actions in a random fashion. The UCB selects an arm a_t at any given time according to the following equation [30]

$$A_t = \underset{a}{arg\,max} \left[Q_t(a) + \sqrt{\frac{c \ln t}{N_t(a)}} \right], \tag{28}$$

where c is the degree of exploration, t is the time step, and N_t is the number of times that the arm a has been selected. The square root part in (28) acts as the variance of the estimated value of action a and it shows the level of uncertainty about the action. When an action is selected, the N_t for that action is increased, since this term resides in the denominator of (28), the whole term under the square root decreases. However, when other actions are selected, the value of t in the nominator increases, therefore, the uncertainty increases. This increment in the uncertainty is logarithmic, which means that the value of this increment will get smaller by time. This will guarantee that the actions that have a lower estimate value or that have been selected for a large number of times will not be selected frequently in the future. The UCB algorithm is summarized in Algorithm 2 [53].

One way to measure the performance of an MAB algorithm is by calculating its regret. The regret is the difference in reward of the best possible arm and the reward of the arm that was played. In order to compute the regret we assume that we know the probability distribution from which each action is selected, therefore, we can pick the optimal action by choosing the action with the highest payoff. The regret calculation can be written as [53]

$$L_t = T\mathbb{E}[R_t|A_t = a^*] - \sum_t \mathbb{E}[R_t|A_t = a_t], \tag{29}$$

where a^* is the optimal action given the probability distributions of all the actions and it can be found by

$$a^* = \max_{a \in A} \mathbb{E}[R_t|A_t = a_t], \tag{30}$$

4.3. Problem Formulation

In this section, we formulate the problem of relay positioning as a MAB problem. In our formulation, There is a maximum number of l predefined locations for the relay to accommodate.

4.3.1. Problem formulation

Consider the matrix of the locations for the relay defined in (13). Let $l_j \in \mathbb{R}^3$ be the ith location that the relay can operate at, where the first, second, and the third element of l_j are x, y, and z coordinates, respectively. Each location for the relay is considered as the arm of a bandit machine. We refer to these arms as the actions and show them by $a \in A = \{a_1, a_2, ..., a_l\}$. The relay will establish two links regardless of the location that it operates at. One of the links is from the v to r and the other one from the r to b. Each of these links will have a rate which determines if the link is proper or not. Therefore, for a given coordinate for the v on the ground we can calculate the value of the rates for each of the locations and store them in a vector. Each element in this vector of rates is considered to be the reward r_t assigned to the locations of the relay.

In order to find the rates associated with each relay location we define two vectors $\boldsymbol{s}_{vr}, \in \mathbb{R}^{1\times l}$ and $\boldsymbol{s}_{rb} \in \mathbb{R}^{1\times l}$ which contain the received powers at each location of r and the received powers in b from each location of the relay. The vector $\boldsymbol{s}_{vr}$ can be expressed as:

$$\boldsymbol{s}_{vr} = p_v \boldsymbol{g}_{vr}. \tag{31}$$

where p_v is the transmit power of the v and $\boldsymbol{g}_{vr} \in \mathbb{R}^L$ is channel gain vector for the links between v and each of the predefined locations for r. Similarly, the vector $\boldsymbol{s}_{rb}$ is given as:

$$\boldsymbol{s}_{rb} = p_r \boldsymbol{g}_{rb}. \tag{32}$$

where p_r is the transmit power of the r and $\boldsymbol{g}_{rb} \in \mathbb{R}^L$ is vector of the channel gains for the links between each of the predefined locations of r and b.

By using (31) and (32) the SNRs of the links can be written in the form of vectors. The SNR vector for the links between the v and r can be given as:

$$\boldsymbol{\gamma}_{vr} = \frac{\boldsymbol{s}_{vr}}{N_0}. \tag{33}$$

Similarly, the SNR vector for the links between the different locations of r and b can be given as:

$$\boldsymbol{\gamma}_{rb} = \frac{\boldsymbol{s}_{rb}}{N_0}. \tag{34}$$

Now we can calculate the rate for each link using the calculated SINR and the SNR calculated above

$$\boldsymbol{r}_{sr} = \log_2(\boldsymbol{\gamma}_{sr} + 1), \tag{35}$$

$$\boldsymbol{r}_{rb} = \log_2(\boldsymbol{\gamma}_{rb} + 1), \tag{36}$$

$$\boldsymbol{r}_t = \boldsymbol{r}_{sr} + \boldsymbol{r}_{rb}, \tag{37}$$

where $\boldsymbol{r}_{sr}$ is the vector of the rates between the s and all the l locations of the relay. Similarly, the $\boldsymbol{r}_{rb}$ is the vector of the rates for the links between all the possible locations of the relay and b. Moreover, $\boldsymbol{r}_t$ is a vector including the total rate for each location of the relay. The total rate is calculated by adding up the rates of the uplink and the downlink.

The goal of the relay is to find the location with the maximum sum rate. Since the MAB framework is designed to learn how to act in one specific situation, we play this game only for one particular given source node location and find the proper location for the UAV which can provide the best rate.

In the bandit problem, each time an action is selected and the reward for that action is selected from $\boldsymbol{r}_t$. The objective of the relay is to maximize the rewards that it attains by selecting the location which provides the maximum rate for the given coordination of the source node on the ground.

4.4. Numerical Results

We consider a cross-road in which the vehicle is located. The pre-defined locations for the relay are considered to be above this cross-road. These locations are in a square area of 226 m $\times$ 226 m with 32 m distance between them which make up for 64 locations in total. We consider the communications to be the carrier frequency of 2 GHz and the parameters used to calculate the air to ground channel for different environments are given in Table 1 [32]. Similar to section 3.3 we assume the BS to be at the coordination of (1000,1000) and the location of the source node to be selected randomly along the streets of the length 1 Km.

Figure 15 shows the effect of the epsilon in the epsilon-greedy algorithm. With an epsilon of zero, the algorithm acts in a pure greedy manner. It locks its selection on the first arm that it plays and it only selects that arm in the future. Since there is no exploration of the other options it does not have any idea if there are any better actions. By selecting a non-zero epsilon we introduce the concept of exploration to the scenario. In this case, the algorithm starts selecting the sub-optimal arms randomly, and the number of random selections depends on the value of the epsilon. Moreover, it keeps the track of the rewards and the number of times that an arm is played to calculate an estimate of the action value for each arm. Therefore, as time goes by, the algorithm can exploit the action with high value and explore to find other actions which might be better than the one that currently is being exploited. However, with an epsilon of 1, the algorithm is a pure exploring machine. It acts like a naive algorithm, where, it chooses the arms in a random way without any knowledge about the estimates of the action values.

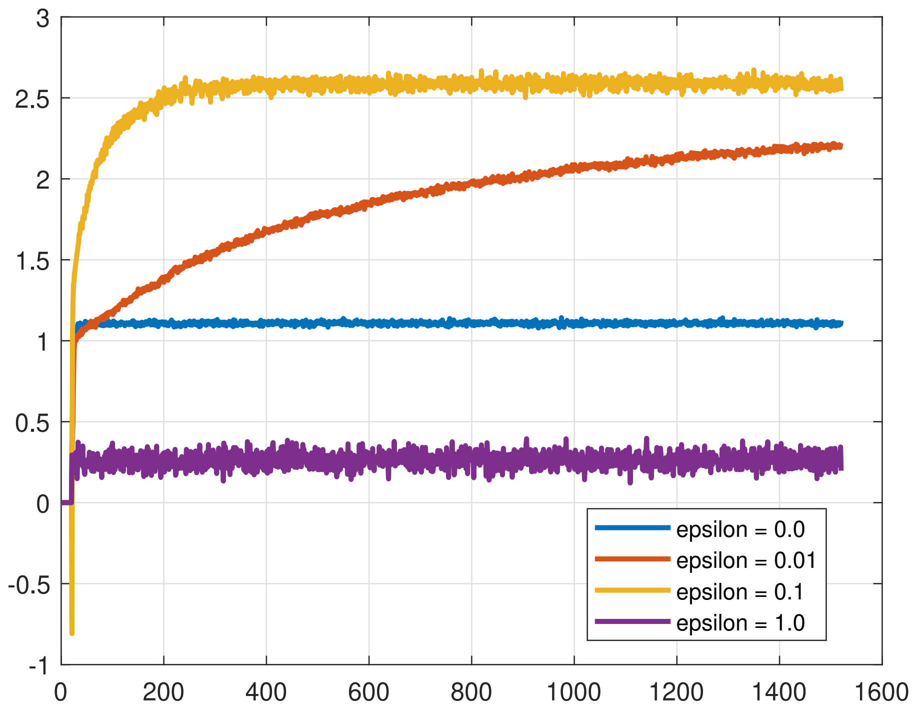

Figure 15: Total average reward for greedy algorithms.

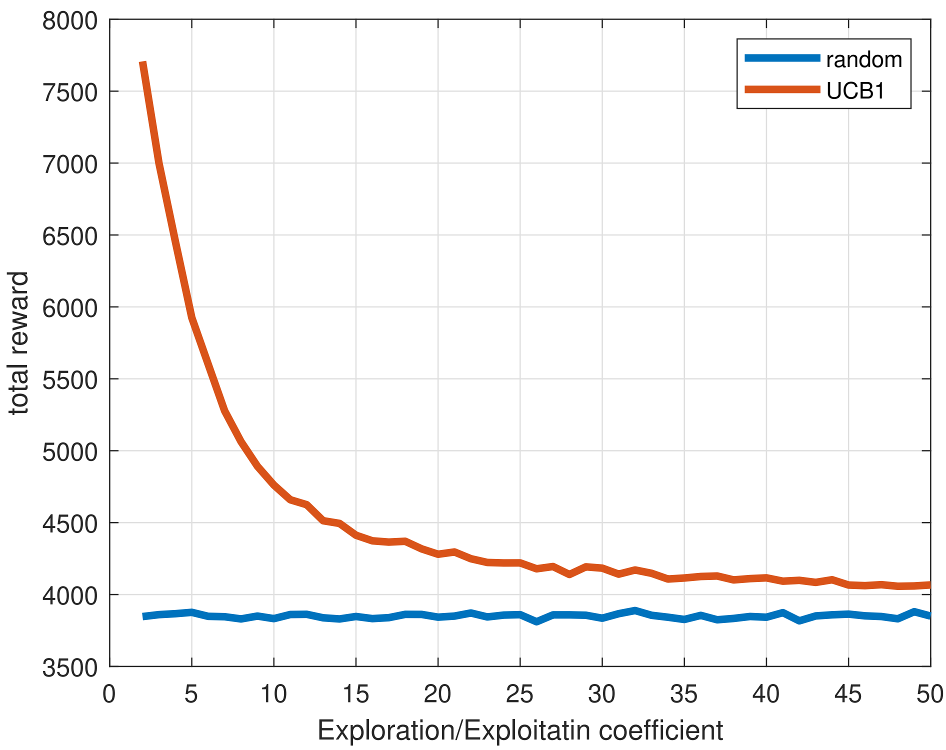

Figure 16: behavior of total cumulative reward by changing the exploration coefficient.

Figure 16 shows the effect of exploration coefficient in the UCB algorithm. The red curve indicates the total reward gained from the UCB algorithm whereas the blue curve is that of the naive approach. The x-axis is the explore-exploit coefficient used in the algorithm, and the y-axis is the amount of total reward. By increasing the exploration coefficient the total reward of the UCB algorithm decreases. This decrement happens due to increment in exploration which leads to decrement in the amount of exploitation of the optimal action. Therefore, the accumulative reward after a finite set of selections reduces. This parameter does not have any effects on the rewards gained from the naive algorithm because this algorithm is based on total exploration.

Figure 17 shows the regret for three algorithms used to solve the relay positioning problem. The red curve shows the regret of the UCB algorithm and as expected it is a logarithmic curve. The blue curve is regret associated with the naive algorithm and the yellow one is the regret produced by the ϵ-greedy algorithm. The regrets of the naive and the ϵ-greedy algorithms increase linearly. This behavior shows the effect of knowledge about the probability distributions from which the rewards are drawn. The regret for the naive algorithm increases more rapid than the other two because it has no information about the optimal action. However, the ϵ-greedy algorithm does better compare to the naive one, because in addition to the exploitation of the best action till that point in time it explores the unknown action and by keeping the record of the previous selections it makes its decisions for the future selections. The reason for the linear regret is that these algorithms select the sub-optimal arms very often.

On the other hand, the UCB algorithm minimizes regret by identifying an optimal arm and playing it. It finds the optimal arm by collecting data from the previous rounds and computing an estimate of the action values for each arm and eliminating the arms that seem to be sub-optimal.

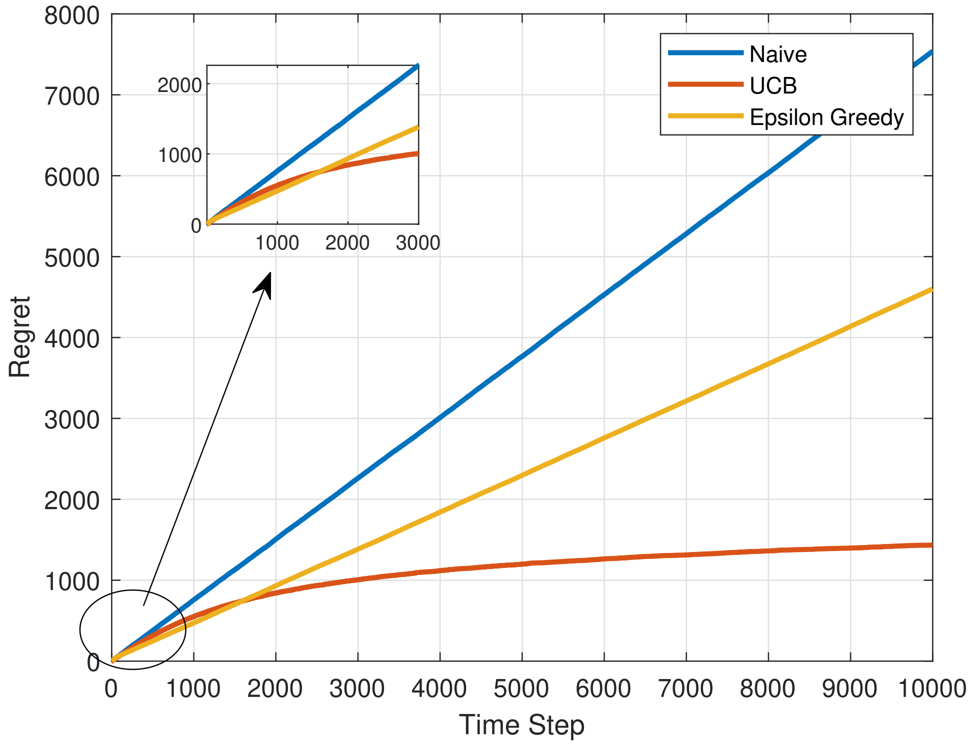

Figure 17: Cumulative regret for UCB and naive algorithm, hight of UAV=50m.

Figure 18 and Figure 19 show the normalized reward that we gain at each time step when we use the UCB algorithm and the naive algorithm respectively. In Figure 18, at the first time steps the algorithm chooses the locations arbitrary because it does not have any knowledge about the reward distribution of each of the locations, however, as time passes the algorithm starts to gain awareness about the distributions and picks the location for the relay based on its knowledge which action would yield a better reward. That is why towards the end of the x-axis the rewards tend to be mostly high. However, in the case of the naive algorithm, there is no difference between the first time steps and the last time steps because the naive algorithm keeps choosing the locations randomly.

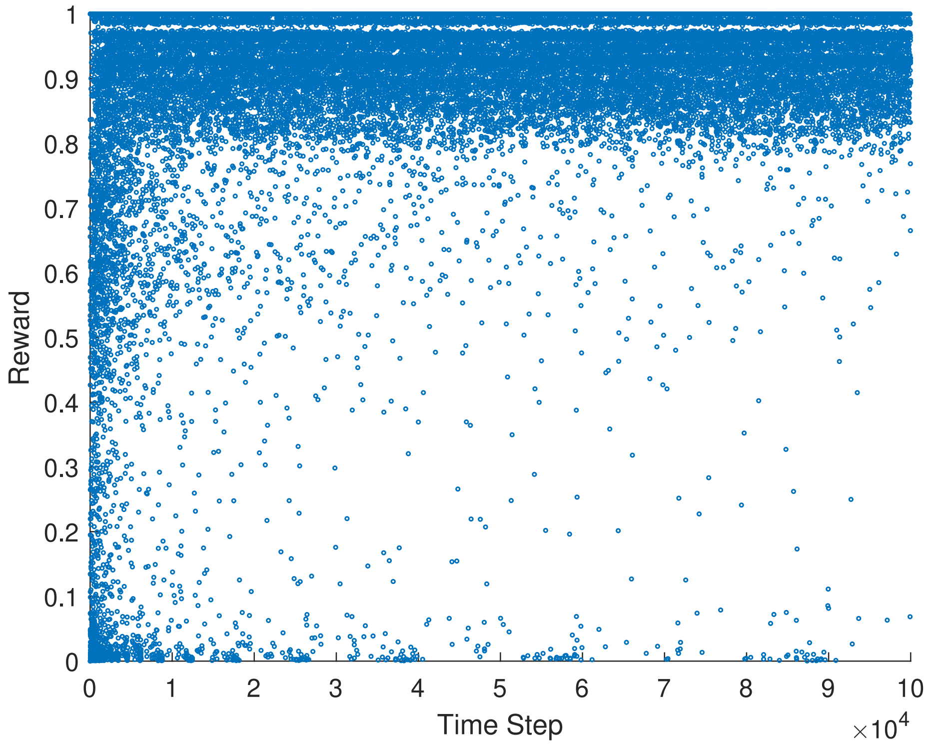

Figure 18: Random rewards gained at each time step for the UCB algorithm.

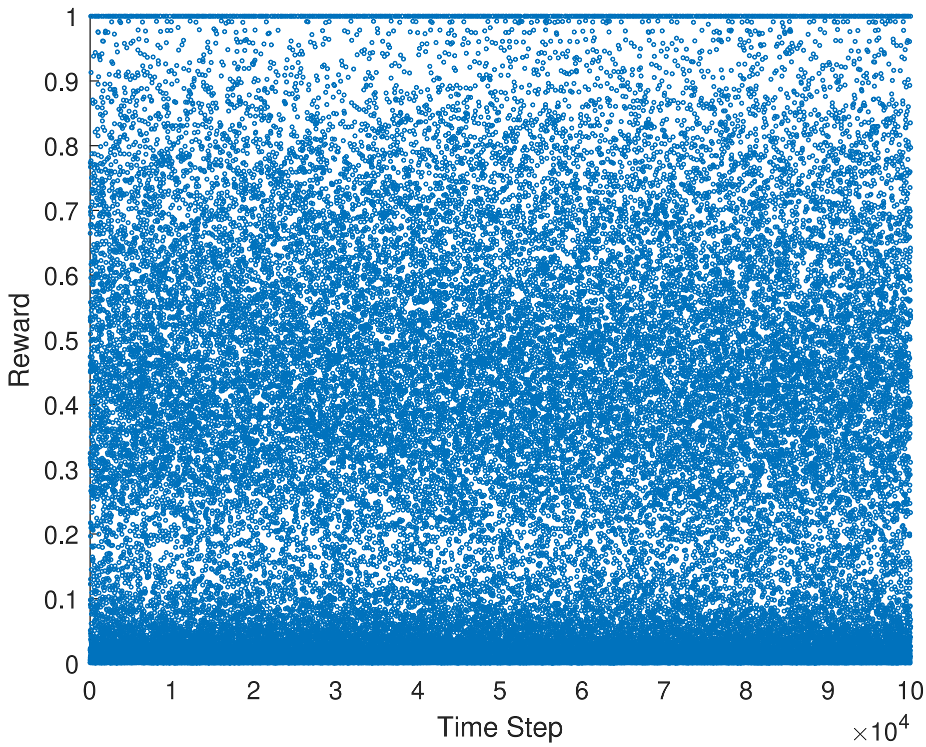

Figure 19: Random rewards gained at each time step for the naive algorithm.

Fig. 20 shows the cumulative reward for each of the locations. All of the sub-optimal locations would yield low rates and the UCB algorithm should not select them. The highest cumulative reward which is highlighted in Fig. 20 belongs to the optimal location. The UCB algorithm successfully identifies that arm and selects it more often, which leads to higher total throughput in the system.

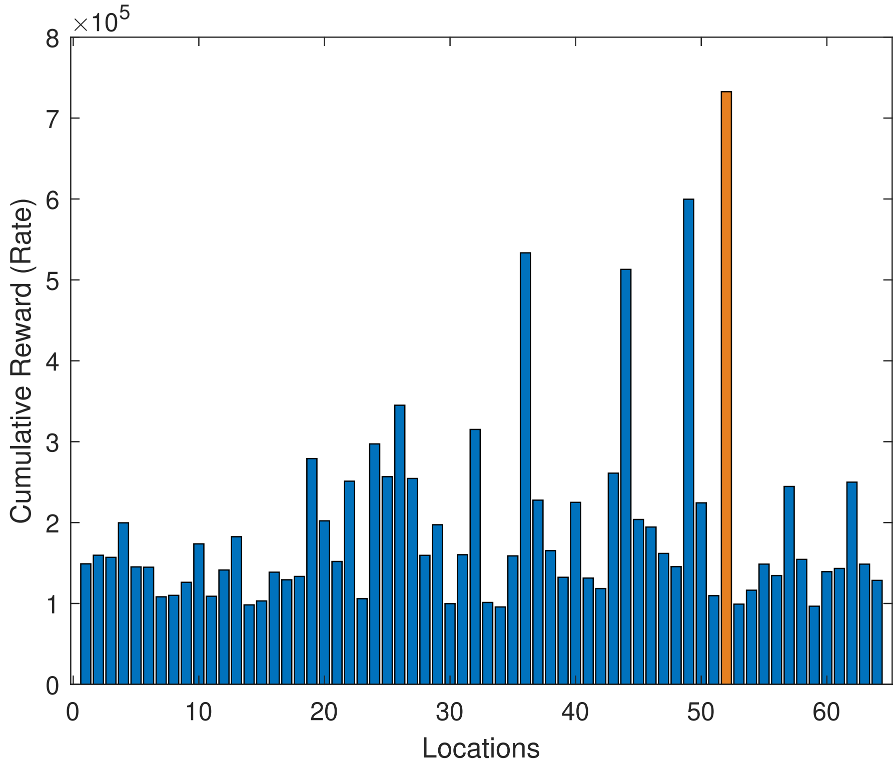

Figure 20: The cumulative reward for each location.

Figure 21 shows the upper confidence bound after the UCB algorithm is done running and Figure 22 depicts the number of the selections for each action. The action number 32 which corresponds to the 32nd location for the relay has the highest number of selections with 15,320 times out of 20,000. Moreover, this in Figure 21 the confidence bound at the 32nd index has the lowest value. As we expected the action with the highest number of selections has the lowest confidence bound. The reason is that in Equation 27, the term under the square root which indicates the confidence bound keeps decreasing as the number of the times that an action is selected increases. On the other hand, when an action is not selected, the number of times which that action has been selected is not altered but the nominator of the square root is increasing. Therefore, all the actions other than the optimal action have higher confidence bounds.

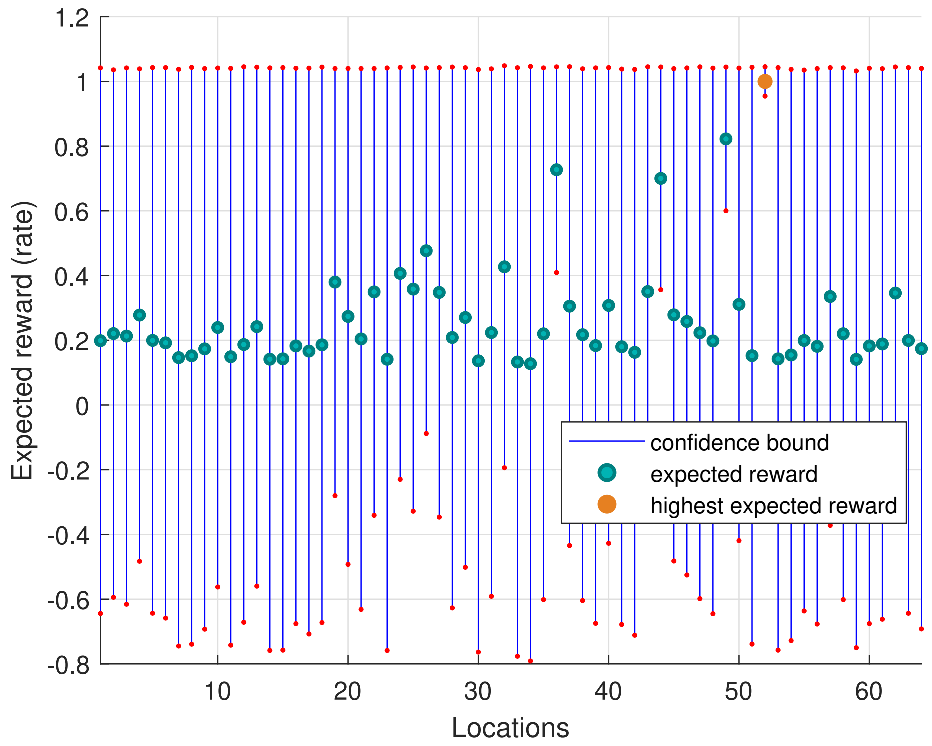

Figure 21: The normalized expected reward of each location andthe confidence bound.

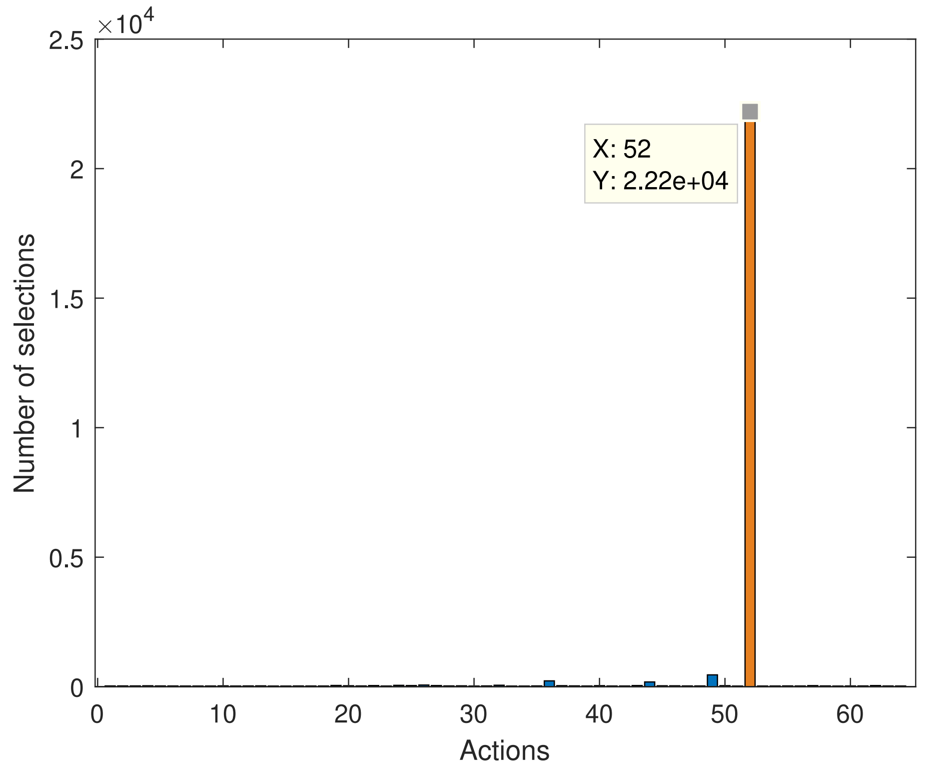

Figure 22: The number of times each action is selected using the UCB algorithm.

5. CONCLUSION

In this work, FD UAV relaying is proposed to increase wireless coverage in vehicular communication networks with an underlay V2V link. First, by using a set of predefined locations for the UAV relay, and, also by considering the locations of the vehicles on the ground, we have derived the SINRs for all the possible locations for the UAV. Second, to find the optimal location of the UAV, we have formulated an ℓ_0-norm minimization problem. Finally, since the formulated problem is non-combinatorial and NP-hard, we have used an ℓ_1-norm approximation for it, which results in a convex optimization problem. Simulation results have shown that by using the proposed method, the number of times that the UAV can find a location to satisfy the SINR requirements of all the links is 10% higher compared a baseline scenario in which the UAV has a fixed location.

Moreover, the problem of FD UAV positioning is solved using the multi-armed bandit framework of reinforcement learning. In this approach, the predefined UAV locations are considered as the actions of the MAB. The objective of the UAV is to select the location where it can maximize the total sum rate of the network. The performances of different algorithms such as naive, greedy, ϵ-greedy, and UCB are studied and shown that the UCB algorithm performs better than others by achieving a logarithmic regret. Simulation results show that the UCB algorithm can successfully find the desired location for the UAV which would yield the highest sum rate among all the predefined locations.

6. REFERENCES

[1] Goldsmith A. (2005) Wireless communications. Cambridge university press.

[2] Tse D. & Viswanath P. (2005) Fundamentals of wireless communication. Cambridge university press.

[3] Mietzner J., Schober R., Lampe L.H.J., Gerstacker W.H., Hoeher P.A. et al. (2009) Multiple-antenna techniques for wireless communications-a comprehensive literature survey. IEEE Communications Surveys and Tutorials 11, pp. 87–105.

[4] Wu C.C., Lee W.B. & Tsaur W.J. (2008) A secure authentication scheme with anonymity for wireless communications. IEEE Communications Letters 12, pp. 722–723.

[5] Sharef B.T., Alsaqour R.A. & Ismail M. (2014) Vehicular communication ad hoc routing protocols: A survey. Journal of network and computer applications 40, pp. 363–396.

[6] Al-Sultan S., Al-Doori M.M., Al-Bayatti A.H. & Zedan H. (2014) A comprehensive survey on vehicular ad hoc network. Journal of network and computer applications 37, pp. 380–392.

[7] Choi J., Va V., Gonzalez-Prelcic N., Daniels R., Bhat C.R. & Heath R.W. (2016) Millimeter-wave vehicular communication to support massive automotive sensing. IEEE Communications Magazine 54, pp. 160–167.

[8] Mozaffari M., Saad W., Bennis M. & Debbah M. (2017) Mobile unmanned aerial vehicles (uavs) for energy-efficient internet of things communications. IEEE Transactions on Wireless Communications 16, pp. 7574–7589.

[9] Xue Z., Wang J., Ding G. & Wu Q. (2018) Joint 3d location and power optimization for uav-enabled relaying systems. IEEE Access 6, pp. 43113–43124.

[10] Agiwal M., Roy A. & Saxena N. (2016) Next generation 5g wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials 18, pp. 1617–1655.

[11] Gupta A. & Jha R.K. (2015) A survey of 5g network: Architecture and emerging technologies. IEEE access 3, pp. 1206–1232.

[12] Yang N., Wang L., Geraci G., Elkashlan M., Yuan J. & Di Renzo M. (2015) Safeguarding 5g wireless communication networks using physical layer security. IEEE Communications Magazine 53, pp. 20–27.

[13] Gubbi J., Buyya R., Marusic S. & Palaniswami M. (2013) Internet of things (iot): A vision, architectural elements, and future directions. Future generation computer systems 29, pp. 1645–1660.

[14] Kortuem G., Kawsar F., Sundramoorthy V. & Fitton D. (2010) Smart objects as building blocks for the internet of things. IEEE Internet Computing 14, pp. 44–51.

[15] Whitmore A., Agarwal A. & Da Xu L. (2015) The internet of things a survey of topics and trends. Information Systems Frontiers 17, pp. 261–274.

[16] Da Xu L., He W. & Li S. (2014) Internet of things in industries: A survey. IEEE Transactions on industrial informatics 10, pp. 2233–2243.

[17] Bharadia D., McMilin E. & Katti S. (2013) Full duplex radios. In: ACM SIGCOMM Computer Communication Review, vol. 43, ACM, vol. 43, pp. 375–386.

[18] Choi J.I., Jain M., Srinivasan K., Levis P. & Katti S. (2010) Achieving single channel, full duplex wireless communication. In: Proceedings of the sixteenth annual international conference on Mobile computing and networking, ACM, pp. 1–12.

[19] Sabharwal A., Schniter P., Guo D., Bliss D.W., Rangarajan S. & Wichman R. (2014) In-band full-duplex wireless: Challenges and opportunities. IEEE Journal on Selected Areas in Communications 32, pp. 1637–1652.

[20] Aryafar E., Khojastepour M.A., Sundaresan K., Rangarajan S. & Chiang M. (2012) Midu: Enabling mimo full duplex. In: Proceedings of the 18th annual international conference on Mobile computing and networking, ACM, pp. 257–268.

[21] Knox M.E. (2012) Single antenna full duplex communications using a common carrier. In: Wireless and microwave technology conference (WAMICON), 2012 IEEE 13th annual, IEEE, pp. 1–6.

[22] Everett E., Duarte M., Dick C. & Sabharwal A. (2011) Empowering full-duplex wireless communication by exploiting directional diversity. In: Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on, IEEE, pp. 2002–2006.

[23] Bhoi S.K. & Khilar P.M. (2014) Vehicular communication: a survey. IET Networks 3, pp. 204–217.

[24] Hartenstein H. & Laberteaux L. (2008) A tutorial survey on vehicular ad hoc networks. IEEE Communications magazine 46, pp. 164–171.

[25] Araniti G., Campolo C., Condoluci M., Iera A. & Molinaro A. (2013) Lte for vehicular networking: a survey. IEEE communications magazine 51, pp. 148–157.

[26] Mozaffari M., Saad W., Bennis M., Nam Y.H. & Debbah M. (2018) A tutorial on uavs for wireless networks: Applications, challenges, and open problems. arXiv preprint arXiv:1803.00680 .

[27] Tang X. & Hua Y. (2007) Optimal design of non-regenerative mimo wireless relays .

[28] Chandra A., Bose C. & Bose M.K. (2011) Wireless relays for next generation broadband networks. IEEE Potentials 30, pp. 39–43.

[29] Goodfellow I., Bengio Y., Courville A. & Bengio Y. (2016) Deep learning, vol. 1. MIT press Cambridge.

[30] Sutton R.S. & Barto A.G. (2018) Reinforcement learning: An introduction. MIT press.

[31] Belanovic P., Valerio D., Paier A., Zemen T., Ricciato F. & Mecklenbrauker C.F. (2010) On wireless links for vehicle-to-infrastructure communications. IEEE Transactions on Vehicular Technology 59, pp. 269–282.

[32] Al-Hourani A., Kandeepan S. & Jamalipour A. (2014) Modeling air-to-ground path loss for low altitude platforms in urban environments. In: 2014 IEEE Global Communications Conference, pp. 2898–2904.

[33] Mozaffari M., Saad W., Bennis M. & Debbah M. (2016) Mobile internet of things: Can uavs provide an energy-efficient mobile architecture? In: 2016 IEEE Global Communications Conference (GLOBECOM), pp. 1–6.

[34] Cai Y., Yu F.R., Li J., Zhou Y. & Lamont L. (2013) Medium access control for unmanned aerial vehicle (UAV) ad-hoc networks with full-duplex radios and multipacket reception capability. IEEE Transactions on Vehicular Technology 62, pp. 390–394.

[35] Mozaffari M., Saad W., Bennis M. & Debbah M. (2016) Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs. IEEE Transactions on Wireless Communications 15, pp. 3949–3963.

[36] Bazzi A., Masini B.M., Zanella A. & Thibault I. (2017) On the performance of IEEE 802.11 p and LTE-V2V for the cooperative awareness of connected vehicles. IEEE Transactions on Vehicular Technology 66, pp. 10419–10432.

[37] Coll-Perales B., Gruteser M. & Gozalvez J. (2018) Evaluation of IEEE 802.11 ad for mmWave V2V communications. In: Wireless Communications and Networking Conference Workshops (WCNCW), 2018 IEEE, IEEE, pp. 290–295.

[38] Campolo C., Molinaro A., Berthet A.O. & Vinel A. (2017) Full-duplex radios for vehicular communications. IEEE Communications Magazine 55, pp. 182–189.

[39] Ali S., Rajatheva N. & Latva-aho M. (2014) Full duplex device-to-device communication in cellular networks. In: 2014 European Conference on Networks and Communications (EuCNC), pp. 1–5.

[40] Zeng F., Zhang R., Cheng X. & Yang L. (2018) UAV-assisted data dissemination scheduling in VANETs. In: 2018 IEEE International Conference on Communications (ICC), pp. 1–6.

[41] Muruganathan S.D., Lin X., Maattanen H.L., Zou Z., Hapsari W.A. & Yasukawa S. (2018) An overview of 3gpp release-15 study on enhanced lte support for connected drones. arXiv preprint arXiv:1805.00826 .

[42] Wang X., Fu L., Zhang Y., Gan X. & Wang X. (2016) Vdnet: an infrastructure-less uav-assisted sparse vanet system with vehicle location prediction. Wireless Communications and Mobile Computing 16, pp. 2991–3003.

[43] Abbas T., Sjöberg K., Karedal J. & Tufvesson F. (2015) A measurement based shadow fading model for vehicle-to-vehicle network simulations. International Journal of Antennas and Propagation 2015.

[44] Al-Hourani A., Kandeepan S. & Lardner S. (2014) Optimal lap altitude for maximum coverage. IEEE Wireless Communications Letters 3, pp. 569–572.

[45] Mozaffari M., Saad W., Bennis M. & Debbah M. (2015) Drone small cells in the clouds: Design, deployment and performance analysis. In: 2015 IEEE Global Communications Conference (GLOBECOM), pp. 1–6.

[46] Boyd S. (2007) ℓ_1-norm methods for convex-cardinality problems. Available: `http://www.stanford.edu/class/ee364b/lectures/l1_slides.pdf`.

[47] Jiang C., Zhang H., Ren Y., Han Z., Chen K.C. & Hanzo L. (2017) Machine learning paradigms for next-generation wireless networks. IEEE Wireless Communications 24, pp. 98–105.

[48] Ye H., Li Y.G. & Juang B.H.F. (2019) Deep reinforcement learning for resource allocation in v2v communications. IEEE Transactions on Vehicular Technology .

[49] Mastronarde N. & van der Schaar M. (2011) Fast reinforcement learning for energy-efficient wireless communication. IEEE Transactions on Signal Processing 59, pp. 6262–6266.

[50] Xiao L., Lu X., Xu D., Tang Y., Wang L. & Zhuang W. (2018) Uav relay in vanets against smart jamming with reinforcement learning. IEEE Transactions on Vehicular Technology 67, pp. 4087–4097.

[51] Zhang B., Mao Z., Liu W. & Liu J. (2015) Geometric reinforcement learning for path planning of uavs. Journal of Intelligent & Robotic Systems 77, pp. 391–409.

[52] Challita U., Saad W. & Bettstetter C. (2018) Deep reinforcement learning for interference-aware path planning of cellular-connected uavs. In: 2018 IEEE International Conference on Communications (ICC), IEEE, pp. 1–7.

[53] Auer P., Cesa-Bianchi N. & Fischer P. (2002) Finite-time analysis of the multi-armed bandit problem. Machine learning 47, pp. 235–256.